



University
of Glasgow

Martinez Corrales, Guillermo (2019) *Identification of novel transcription factors in the development and function of the Drosophila renal tubule.*

PhD thesis.

<https://theses.gla.ac.uk/74261/>

Copyright and moral rights for this work are retained by the author

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge

This work cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given

Enlighten: Theses

<https://theses.gla.ac.uk/>
research-enlighten@glasgow.ac.uk



University
of Glasgow

**Identification of novel transcription factors in the
development and function of the *Drosophila* renal
tubule**

by

Guillermo Martínez Corrales

Submitted for the degree of
Doctor of Philosophy at the University of Glasgow

Institute of Molecular, Cell and Systems Biology
College of Medical, Veterinary and Life Sciences

University of Glasgow

Glasgow

G12 8QQ

May 2019

Abstract

The formation and development of an organ is a complex process which involves several programmed events, including changes in cell shape and adhesion, proliferation, and differentiation. Defects in any of them lead to malformations and lethality. Investigating these processes can lead to mechanistic understanding of organ development in health and disease conditions. They are defined by the expression and combination of specific Transcription Factors (TFs). These are active during the development of the Malpighian Tubules (MTs), of *Drosophila melanogaster*, which perform equivalent functions to the human kidney. Therefore, research using *Drosophila* proved very useful for understanding processes of cell differentiation, proliferation, and development.

To investigate which novel TFs could be involved in the development and physiology of the *Drosophila* kidney, an initial genetic screen of candidate TFs was performed. This selection criteria included the identification of candidate TFs regarding their patterns of expression, previous literature and human homologues involved in renal diseases. This resulted in the selection *GATAe*, a member of the GATA family of TFs, as the focus of this project.

The effects of *GATAe* downregulation and overexpression were studied in the context of the MTs employing diverse genetic manipulation techniques. Silencing *GATAe* in any of the three main cell-types of the MTs occasioned a diverse range of phenotypes.

GATAe in the MT Principal Cells (PCs) is not required for the embryonic development of the tubule, but is crucial in to ensure the correct morphology and maintenance of the adult tubule from metamorphosis. As a consequence, loss of *GATAe* in PCs results in a reduced lifespan, less tolerance to diverse kinds of stress, and liquid retention. Further preliminary ChIP-Sequencing analysis using a tagged line of *GATAe* revealed potential downstream targets of *GATAe* in the PCs, reinforcing its roles in cell-to-cell communication, tissue maintenance and stress response in the MTs.

In Stellate Cells (SCs), *GATAe* is required for their survival through metamorphosis. Therefore, loss of *GATAe* in SCs induces a sharp reduction in this

cell population in the adult stage, and impairing their secretion capabilities. Furthermore, *GATAe* is necessary in the Renal and Nephritic Stem Cells (RNSCs) for their migration into the ureters and lower tubules during metamorphosis.

Overall, the findings reported in this PhD thesis reveal the novel functions of *GATAe* as a potential master regulator in the events of tissue maintenance, cell survival and cell migration. Apart from *GATAe*, additional TFs were identified in this project, which could serve as a base for future investigation of their possible novel roles in the fly renal system. Altogether, the data showed here provides new insights into the molecular pathways involved in the formation and maintenance of a functional organ.

Table of contents

Abstract	2
Table of contents	4
List of Tables	9
List of Figures	10
Acknowledgements	13
Author's Declaration	14
Abbreviations	15
Chapter 1 - Introduction	17
1.1 <i>Drosophila melanogaster</i> as a model system	18
1.1.1 Overview	18
1.1.2 <i>Drosophila</i> genome	19
1.1.3 <i>Drosophila</i> genetic tools	20
1.1.4 <i>Drosophila</i> life cycle	24
1.2 The <i>Drosophila</i> Malpighian tubule	25
1.2.1 Overview	25
1.2.2 Structure and function of the MTs	26
1.2.3 Development of the MTs	28
1.2.4 Transcriptomic profile of the adult Malpighian tubules	34
1.3 Transcription Factors	36
1.3.1 Overview	36
1.3.2 Transcription factors, development and disease	37
1.3.3 Conservation of transcription factor binding specificities across evolution	37
1.3.4 Transcription factors enriched in the <i>Drosophila</i> MTs	38
1.4 The GATA family	39
1.4.1 Gene family and structure	39
1.4.2 GATA factors in vertebrates	40
1.4.3 GATA factors in <i>Drosophila</i>	45
1.4.4 <i>GATAe</i>	46
1.5 Aims of the project	51
Chapter 2 - Materials and Methods	52
2.1 <i>Drosophila melanogaster</i>	53
2.1.1 Rearing conditions	53
2.1.2 Fly stocks	53
2.1.3 Embryo collection	56
2.1.4 Dissection of <i>Drosophila</i> MTs	56

2.2	<i>Escherichia coli</i>	56
2.2.1	Strains, Vectors and Transformation conditions	56
2.2.2	Isolation of plasmid DNA.....	58
2.2.3	Verification of plasmid insertion.....	58
2.3	<i>Drosophila</i> S2 Cells	59
2.3.1	Maintenance	59
2.3.2	Transient transfection.....	59
2.4	Molecular biology protocols	60
2.4.1	RNA extraction	60
2.4.2	cDNA synthesis.....	60
2.4.3	Polymerase chain reaction (PCR)	61
2.4.4	Restriction enzyme digestion.....	62
2.4.5	Agarose gel electrophoresis of DNA	63
2.4.6	Western blotting	63
2.5	Generation of the antibody against GATAe protein	65
2.5.1	Antigenic peptide design	65
2.6	Immunofluorescence	65
2.6.1	Immunofluorescence of larval, pupal and adult tubules	65
2.6.2	Immunofluorescence of Embryos	66
2.6.3	Immunofluorescence of S2 cells	66
2.6.4	List of primary and secondary antibodies used	67
2.6.5	Whole fly imaging sample preparation.....	69
2.6.6	Larvae preparation.....	69
2.7	Imaging.....	69
2.7.1	Confocal Microscopy.....	69
2.7.2	Optical Tomography Projection Microscopy	70
2.7.3	Two-photon Microscopy.....	70
2.8	Chromatin immunoprecipitation sequencing.....	71
2.8.1	Sample collection	72
2.8.2	Nuclei preparation	73
2.8.3	Chromatin immunoprecipitation	74
2.8.4	Chromatin elution and reversal crosslink	75
2.8.5	DNA purification.....	75
2.8.6	ChIP-sequencing.....	75
2.9	Physiology experiments	76
2.9.1	Lifespan assay	76
2.9.2	Starvation and desiccation assays.....	76
2.9.3	Gravimetric estimations of body water	77

2.9.4	Ramsay secretion assay	77
2.10	Bioinformatics	79
Chapter 3 - Genetic Screen of novel Transcription Factors involved in Malpighian Tubule Development and Morphology		80
3.1	Summary	81
3.2	Results	81
3.2.1	List of most enriched tubule TFs	81
3.2.2	TFs enrichment in adult MTs	81
3.2.3	TFs expression during embryogenesis	82
3.2.4	Human TFs orthologues	83
3.2.5	RNA interference lethality screen	83
3.2.6	RNA interference and specific gene silencing in the MTs	84
3.3	Discussion	86
3.3.1	Other TFs to be investigated	88
Chapter 4 - Analysis of <i>GATAe</i> in the Principal Cells of the MTs		90
4.1	Introduction	91
4.2	Morphological characterisation	91
4.2.1	<i>GATAe</i> is expressed in all MT cell types	91
4.2.2	Reduced levels of <i>GATAe</i> lead to morphological abnormalities of the MTs	95
4.2.3	<i>GATAe</i> is not essential during the embryonic development of the MTs	101
4.2.4	<i>GATAe</i> is required during metamorphosis	103
4.2.5	Adult MTs maintenance requires the function of <i>GATAe</i>	105
4.2.6	Loss of <i>GATAe</i> induces cell proliferation	107
4.2.7	Reduced levels of <i>GATAe</i> modulate cancer-related gene expression	109
4.2.8	Knockdown of <i>GATAe</i> in a small number of PCs is enough to induce abnormal tissue growth	111
4.3	Functional characterisation of <i>GATAe</i>	112
4.4	Discussion	114
4.4.1	Evolutionarily conserved functions of the GATA family	118
Chapter 5 - Analysis of <i>GATAe</i> in the Stellate and Renal and Nephritic Stem Cells		121
5.1	Introduction	122
5.2	<i>GATAe</i> in Stellate Cells	122
5.3	<i>GATAe</i> in the RNSCs	125
5.4	Discussion	127
5.4.1	<i>GATAe</i> and SC population	128
5.4.2	<i>GATAe</i> and RNSC population	129
5.5	Proposed model for <i>GATAe</i> function in the MTs	130
Chapter 6 - Molecular and biochemical tools for <i>GATAe</i> functional characterisation		133

6.1	Introduction	134
6.2	Predicted 3D structure of GATAe protein	134
6.3	Generation of an antibody against GATAe	136
6.3.1	Sequence selection	136
6.3.2	Immunocytochemistry and Western-Blot using GATAe antibody	137
6.4	Generation of a gain-of-function line of GATAe	139
6.4.1	S2 cell analysis	140
6.4.2	Tissue-specific overexpression in <i>Drosophila</i>	142
6.5	Chromatin Immunoprecipitation Analysis of GATAe	145
6.5.1	Chromatin Immunoprecipitation	145
6.5.2	Technical ChIP optimisations	145
6.5.3	Results	146
6.5.4	Immunoprecipitation analysis	146
6.5.5	Chromatin Immunoprecipitation	147
6.6	Discussion and troubleshooting	161
6.6.1	Further experiments to assess GATAe molecular function	162
Chapter 7 -	Summary and future work	166
7.1	Summary	167
7.1.1	Genetic screen	167
7.1.2	GATAe involvement in the MTs	167
7.1.3	Molecular and biochemical tools for GATAe functional characterisation	169
7.1.4	Observation of 3D structures in whole-fly preparations	170
7.2	Future work	170
7.2.1	Study of other genes detected in the screening	171
7.2.2	Further research on GATAe	171
7.3	Final conclusion	173
Appendices	174
Appendix 1	<i>Drosophila</i> media composition	174
Appendix 2	Grape juice agar medium preparation	175
Appendix 3	E. coli growth media recipes	176
Appendix 4	List of primers used	177
Appendix 5	5X TBE buffer recipe	178
Appendix 6	10X PBS and 10X TBS recipes	179
Appendix 7	Solutions 1 and 2 for ECL development	180
Appendix 8	CUBIC solution preparation	181
Appendix 9	<i>Drosophila</i> saline recipe	182
Appendix 10	ChIP-Sequencing gene list	183

Appendix 11	Observation of whole-fly Malpighian tubules using tomography projection	
microscopy	186
7.4	Introduction	187
7.5	Results	187
7.6	Discussion.....	191
Publications	193
References	194

List of Tables

Table 1.2-1. Transcription factors enriched in the adult MTs.	36
Table 2.1-1. List of <i>Drosophila</i> stocks used in this study including the stock name, genotypes, the purpose and the origin of the stocks.	55
Table 2.2-1. Vectors used in this study.	57
Table 2.4-1. Normal PCR thermal cycling conditions.	61
Table 2.4-2. qPCR thermal cycling conditions.	62
Table 2.6-1. List of primary antibodies used in this project.	67
Table 2.6-2. List of secondary antibodies used in this project.	68
Table 2.6-3. List of other staining markers used.	68
Table 3.2-1. Expression of the shortlisted TF in the embryonic MTs, as revealed by BDGP.	82
Table 3.2-2. <i>Drosophila</i> TFs shortlisted and their closest human homologues.	83
Table 3.2-3. Lethality experiments.	84
Table 6.5-1. Final ChIP DNA concentrations of S2 cells	150
Table 6.5-2. ChIP DNA concentrations of adult MTs.	151
Table 6.5-3. Genes involved in transport called as present in the MTs.	157
Table 6.5-4. Table of the most tubule-enriched genes from the ChIP-seq analysis.	161
Table 7.3-1. Recipe for the <i>Drosophila</i> Standard medium (Fly food).	174
Table 7.3-2. Primer sequences and their purpose.	178
Table 7.3-3. Complete gene list detected by the ChIP-Seq.	185

List of Figures

Figure 1.1.1. Schematic diagram of the GAL4 System.	23
Figure 1.1.2. Scheme of the GAL80 ^{ts} system.....	24
Figure 1.1.3. Stages in the life cycle of <i>Drosophila melanogaster</i>	25
Figure 1.2.1. Schematic overview of the function of a PC (left) and a SC (right) of the main segment of the tubule.....	27
Figure 1.2.2. Scheme of the embryonic development of the MTs of <i>Drosophila</i>	32
Figure 1.2.3. Schematic diagram of the adult MTs.	34
Figure 1.4.1. Cladogram of the different GATA factors in vertebrates and invertebrates. .	40
Figure 1.4.2. <i>GATAe</i> genomic location and embryonic pattern of expression.	48
Figure 1.4.3. FlyAtlas2 reveals <i>GATAe</i> expression in L3 and adult stage.....	49
Figure 2.2.1. Schematic map of the pUAST vector, used for <i>GATAe</i> overexpression.....	58
Figure 2.7.1. Schematic diagram of the OPT microscope.	70
Figure 2.7.2. Schematic diagram comparing Two-Photon microscopy vs confocal microscopy.	71
Figure 2.9.1. The Ramsay Secretion Assay.....	78
Figure 3.2.1. Low levels of <i>Muc11A</i> in the PCs affect the integrity of the adult initial segment.....	86
Figure 4.2.1. Schematic representation of the genomic map of <i>GATAe</i> -Gal4 lines.	92
Figure 4.2.2. <i>GATAe</i> >GFP reveals the pattern of expression of <i>GATAe</i>	93
Figure 4.2.3. <i>GATAe</i> > <i>GATAe</i> RNAi pupae are smaller compared to the controls.	94
Figure 4.2.4. ICC of adult <i>GATAe</i> > <i>GATAe</i> RNAi adult MGs	95
Figure 4.2.5. Length comparison of <i>GATAe</i> knockdown MTs.....	96

Figure 4.2.6. MTs with reduced levels of <i>GATAe</i> show reduced cell number	97
Figure 4.2.7. Reduced levels of <i>GATAe</i> in PCs induces strong morphological defects.	98
Figure 4.2.8. Activation of Ras signalling in PCs induces morphological defects.	100
Figure 4.2.9. Silencing <i>GATAe</i> with <i>CapaR-Gal4</i> does not affect MG morphology.	101
Figure 4.2.10. <i>GATAe</i> knockdown embryonic MTs display wild-type morphology.	102
Figure 4.2.11. <i>CapaR>GATAe</i> RNAi MTs present morphological deficiencies from during metamorphosis.	103
Figure 4.2.12. Developmental defects using <i>CapaR^{ts}</i>	104
Figure 4.2.13. <i>GATAe</i> is required for adult maintenance of the MTs.	106
Figure 4.2.14. <i>GATAe</i> knockdown in PCs induces proliferation of RNSCs.	108
Figure 4.2.15. Alternative RNSC and proliferation markers.	109
Figure 4.2.16. Relative mRNA levels of oncogenes and apoptosis-related genes in <i>GATAe</i> knockdown MTs	110
Figure 4.2.17. <i>GATAe</i> knockdown using <i>UrO>GFP</i>	112
Figure 4.3.1 Reducing <i>GATAe</i> expression in PCs affects water homeostasis and stress tolerance.	114
Figure 5.2.1. <i>GATAe</i> downregulation in the SCs induces loss of the SC population and reduced fluid secretion response to kinin stimulation in the adult stage.	123
Figure 5.2.2. SCs show normal localisation in the L3 stage in <i>GATAe</i> knockdown conditions.	124
Figure 5.3.1. RNSC migration to the MTs is impaired upon <i>GATAe</i> knockdown.	126
Figure 5.3.2. Migration defects of RNSCs in <i>GATAe</i> knockdown MTs.	127
Figure 5.5.1 Proposed model for the function of <i>GATAe</i> in the MTs.	132
Figure 6.2.1. Predicted 3D structure, sequence, and domains of <i>GATAe</i> protein.	136

Figure 6.3.1. ICC using GATAe antibody.....	138
Figure 6.3.2. Western blot analysis of GATAe antibody.	139
Figure 6.4.1. Restriction enzyme digest analysis of UAS- <i>GATAe</i> -HA plasmid.....	140
Figure 6.4.2. Nuclear localisation of GATAe-HA in <i>Drosophila</i> S2 cells.	141
Figure 6.4.3. <i>GATAe</i> and putative downstream gene expression in S2 cells transfected with UAS- <i>GATAe</i> -HA.....	142
Figure 6.4.4. Characterisation of <i>CapaR</i> >UAS- <i>GATAe</i> -HA adult MTs.	143
Figure 6.4.5. Overexpression of UAS- <i>GATAe</i> -HA using different Gal4 lines.....	144
Figure 6.5.1. Western-blot analysis of IP experiments for HA tag.....	147
Figure 6.5.2. Micrococcal nuclease treatments of S2 cells.....	149
Figure 6.5.3. Enriched known motifs identified 1kb of the surrounding peaks detected.	152
Figure 6.5.4. <i>De novo</i> motifs enrichment results.....	153
Figure 6.5.5. Classification of ChIP-Seq hits according to their biological processes.	155
Figure 6.5.6. Significant ChIP-Seq detected genes classified according to their molecular function.	159
Figure 6.6.1. Immunocytochemistry of recombinant <i>GATAe</i> to PCs	163
Figure 7.5.1. 3D images of whole adult flies using OPT microscopy.	188
Figure 7.5.2. Adult MTs observed using multiphoton fluorescence microscopy.	190
Figure 7.5.3. 3D structure of embryos observed using two-photon microscopy.....	190

Acknowledgements

I would like to thank Shireen and Julian for their guidance and for giving me the opportunity to be part of their laboratory, and to the RENALTRACT network for providing the support and the great scientific environment in which I was lucky to participate.

I would also like to thank the following people from the Dow/Davies lab: Selim, for his excellent advice and guidance, and for answering every one of my questions, even when they were too many. To Pablo, for his support and expertise, not only in the lab! To Tony for his help and great scientific advice. To Lucy for her proofreading, and Richard for his help with secretion assays.

Also, I would like to thank Anir for the good lunchtime walks, bioinformatics advice and for literally feeding me with part of his lunches (although in this last case, the thanks would not be for him, but Vaishali). Obviously, a big thank you to Suj... sorry, Saurav Ghimire, for his friendship during these three years, and his endless support. Best of luck, see you at the next conference!

A big thank you to Seppo Vainio's Lab, in particular to Veli-Pekka Ronkanien, for his help with the OPT.

Out of the lab, I would like to give a big big thank you to Wenna, for your patience and reading all through my thesis. You are the best!

Last but not least, a big thank you to my family, especially to my parents for all their support during these years, without you, I (obviously) would not be here.

Author's Declaration

The research reported within this thesis is my own work except where otherwise stated and has not been submitted for any other degree. All sources of information used in the preparation of this thesis are indicated by reference.

Guillermo Martínez Corrales

Abbreviations

°C	Degrees Celsius
3D	Three-dimensional
AA	Amino acid
AMG	Anterior midgut
APF	After puparium formation
BDGP	Berkeley <i>Drosophila</i> Genome Project
BDSC	Bloomington <i>Drosophila</i> Stock Centre
BLAST	Basic local alignment search tool
Bp	Base pair
cDNA	Complementary DNA
CDS	Coding sequence
CG	Computed gene
ChIP-Seq	Chromatin immunoprecipitation sequencing
DAPI	4',6-diamidino-2-phenylindole
DBD	DNA binding domain
ddH ₂ O	Double-distilled water
DGRC	<i>Drosophila</i> genomics resource centre
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphate
DSCP	<i>Drosophila</i> synthetic core promoter
DTT	Dithiothreitol
ECL	Enhanced chemiluminescence
ECM	Extracellular matrix
EDTA	Ethylenediaminetetraacetic acid
EGF	Epithelial growth factor
FDR	False discovery rate
FPKM	Fragments Per Kilobase of transcript per Million mapped reads
g	Gram
GAL80.TS	GAL80 thermosensitive
GFP	Green fluorescent protein
h	hours
HA	Human influenza hemagglutinin
HG	Hindgut
HRP	Horseradish peroxidase
ICC	Immunocytochemistry
IGG	Immunoglobulin G
IP	Immunoprecipitation
ISC	Intestinal stem cell
Kb	Kilobases
kDa	Kilodaltons
l	Litre
L1, L2, L3	Larval stage 1, 2 and 3, respectively
MG	Midgut
Min	Minutes
NGS	Normal goat serum
OD ₂₆₀	Optical density unit
OPT	Optical Tomography Projection

ORF	Open reading frame
PANTHER	Protein ANALysis THrough Evolutionary Relationships
PBS	Phosphate buffered saline
PBTA	PBS + Triton + Sodium azide
PC	Principal cell
PCR	Polymerase chain reaction
PFA	Paraformaldehyde
PMG	Posterior midgut
qPCR	Quantitative PCR
RB	Renal blast
RNA	Ribonucleic acid
RNAi	RNA interference
RNSC	Renal and nephritic stem cell
rpm	Revolutions per minute
RT	Room temperature
S2	Schneider 2
SC	Stellate cell
SEM	Standard error of the mean
SSII	Superscript II
TBS	Tris-buffered saline
TF	Transcription factor
THG	Third harmonic generation
TRiP	Transgenic RNAi Project
UAS	Upstream activation sequence
V	Volts
WB	Western-blot
WL3	Wandering L3
μ	Micro

Chapter 1 - Introduction

1.1 *Drosophila melanogaster* as a model system

1.1.1 Overview

Drosophila melanogaster (commonly called “Fruit Fly”, herein referred to as *Drosophila*) has been used in research for more than a century. *Drosophila* was first kept in captivity and utilised in science by Charles Woodworth in Harvard, in the early 1900s, and T.H. Morgan described the first *Drosophila* mutant (*white* eye) (Morgan, 1910). It has been used in a wide range of research fields, from biomedical, genetic and even agricultural approaches, and research with *Drosophila* has contributed to five Nobel prizes (Huang, 2018, Brown, 2017, Van Hiel et al., 2010, Ugur et al., 2016). The success of this organism in research is due to numerous reasons.

D. melanogaster’s life cycle takes around 10 days at 25°C (or 7 days at 29°C). This means that it grows and reproduces much faster than other models, such as mouse or *Xenopus* (Dow and Romero, 2010). In addition, the cost of maintaining a *Drosophila* strain is no more than USD20/year. Also, its complete genome has been sequenced and published (Adams et al., 2000). The *Drosophila* genome is relatively small, and it is comprised of 4 chromosomes. It contains 13,937 coding genes, according to the database Ensembl 78, and almost 70% of them have human homologs. However, if non-coding RNAs and pseudogenes are included in the counting, the *Drosophila* genome contains up to 29,173 different transcripts (Consortium, 2007). The research community has generated mutant lines for over 50% of the fly genes, and they also created inducible knockdown lines for almost all the *Drosophila* genes, which makes this animal a versatile model to perform genetic studies (Dietzl et al., 2007). The wide range of methods to modify gene expression in *Drosophila* allowed scientists to silence or overexpress almost any gene in tissue and time-specific manner to investigate its functions (Brand and Perrimon, 1993, Gratz et al., 2015).

Furthermore, there are numerous databases available related to *Drosophila*, like the online resource FlyBase (Ashburner and Drysdale, 1994). This database consists of a complete repository of genetic data of different species of *Drosophilidae*. Regarding *D. melanogaster*, it includes information on gene expression, gene function, genome mapping, gene homologies, disease models,

among others. Also, FlyBase is linked with other online databases, e.g. Interpro, OrthoDB, or BDGP, includes links to research publications related to the genes of interest, and it is continuously updated. These features make FlyBase the most powerful online databases of *Drosophila* data.

A further key online tool is FlyAtlas, which is a complex transcriptomic study that allows the detection of specific expression patterns and enrichment of genes in the different tissues of *Drosophila* (Chintapalli et al., 2007, Leader et al., 2018).

1.1.2 *Drosophila* genome

The *Drosophila* genome is 180 Mb in size. However, only 120 Mb of it is euchromatin, located mainly in three of its four chromosomes, as the fourth chromosome contains only about 1Mb. The *Drosophila* genome is highly compacted, which, together with its small size, makes this insect a compelling model to be subjected to genome studies. However, not all of the *Drosophila* genes have been studied exhaustively and have yet to be investigated and characterised. There are still a significant number of them that were found by genomic sequencing projects. They were named with the prefix CG (Computed Gene) and a number (e.g. *CG10278*). Once these genes were given a biological function, researchers provided them with a name, and their CG number became a synonym of the new gene symbol. (e.g. the gene *CG10278* was named *GATAe* in 2005, and from then, subsequent studies named it *GATAe*, but *CG10278* still appears in the database as a synonym for *GATAe*).

Drosophila has been employed in a wide range of research fields, from agricultural and pest control (Dow and Davies, 2003), behavioural (Yapici et al., 2008), or developmental and biomedical studies (Wangler et al., 2015). The genetic similarity between the fly and human genomes and the evolutionary conservation of the molecular mechanisms of disease make *Drosophila* an amenable model to study human conditions. Numerous human diseases can be modelled using *Drosophila*, such as Alzheimer's disease (Bouleau and Tricoire, 2015), Parkinson disease (Feany and Bender, 2000), epilepsy (Parker et al., 2011), obesity (Trinh and Boulianne, 2013), cardiovascular disease (Bier and Bodmer, 2004), fragile X syndrome (Chang et al., 2008) and other diseases. In

addition, *Drosophila* has also been widely used for the investigation of diverse kidney diseases by the study of its Malpighian tubules (MTs). The invertebrate MTs play equivalent roles to the human kidney and liver and, accordingly, share numerous genetic and molecular characteristics (Dow and Romero, 2010). Multiple events in the formation of kidney stones have been discovered using *Drosophila* (Hirata et al., 2012, Dow and Romero, 2010), as well as factors involved in renal cancers (Liu et al., 2009, Zeng et al., 2010) or polycystic kidney disease (Gamberi et al., 2017, Millet-Boureima et al., 2018). Furthermore, research in *Drosophila* has been proved as strongly valuable to study other pest insects, making it suitable even to study vector-transmitted diseases, such as malaria or dengue spread by mosquitoes (Schneider and Shahabuddin, 2000, Mukherjee and Hanley, 2010).

1.1.3 *Drosophila* genetic tools

One of the main advantages of *Drosophila* as an animal model is the full range of options available for genetic manipulation. It provides the possibility to alter genetic expression temporally and spatially (Brand and Perrimon, 1993, Caygill and Brand, 2016, Mondal et al., 2007). There are several *Drosophila* stock centres, e.g. the Bloomington *Drosophila* Stock Center (BDSC) or the Vienna *Drosophila* Resource Center (VDRC), among others, that offer numerous mutant and other fly lines that can be used in research. *Drosophila* also has the advantage that any homozygous lethal or sterile mutation can be tracked through the generations in heterozygosis, by the utilisation of balancer chromosomes. These chromosomes are the product of several inversions and mutations which block homologous recombination with their complementary chromosomes, maintaining them stable, viable and heterozygous through generations (Wallace and King, 1951). Numerous types of balancer chromosomes are utilised (in the first, second, third or fourth chromosome), and some of them have been genetically modified to carry different external phenotypic markers which are easily detectable and indicate the presence of these balancers. These markers include Green Fluorescent Protein (GFP), LacZ, among others. Furthermore, the introduction of the recently discovered genome editing technology CRISPR-Cas9 in *Drosophila* (Gratz et al., 2013, Bassett et al., 2013, Du et al., 2018) enhanced the generation of new and more specific mutant

Drosophila strains and had become a widely utilised method to modify specific gene expression.

1.1.3.1 Transposable elements (P-elements)

Mobile transposable elements represent around 10% of the whole *Drosophila* genome and are responsible for multiple mutations, allowing the study of gene function (Georgiev, 1984). Mobile transposable elements from the P transposable element family (P-elements) are among the most used regarding gene engineering, as they are minimal in size (2.9 kb). Their terminal repeated sequence encodes for a transposase, which can be modified to introduce or excise specific gene sequences, such as a wild-type copy of the *white* gene, to identify and follow the integration of these elements (O'Hare and Rubin, 1983, Spradling and Rubin, 1982). Generally, P-elements can be genetically regulated by differential splicing, allowing them to be specifically active in the germline, whereas in the somatic cells they are naturally inactive. Modified P-elements, however, may be adjusted to induce transposition in *Drosophila* embryos if an external transposase is injected together with the desired plasmid.

There is a wide range of different types of P-elements, and among them, the enhancer traps are the most widely characterised. These are modified P-elements that include a reporter gene (such as *LacZ*) instead of the classical transposase gene. Therefore, enhancer traps are not able to induce transcription by themselves, but, if they are introduced in the genome by germline transformation (using an external transposase), they can be inserted into the genome, and, in consequence, their expression is regulated by the nearest gene. Therefore, the natural pattern of expression of a particular gene can be examined utilising these P-elements. The first generation of enhancer traps generated included the *β-Galactosidase* reporter, which could be detected by X-Gal staining (Bier et al., 1989, O'Kane and Gehring, 1987). The BDGP (Berkeley *Drosophila* Genome Project) gene disruption project identified more than 30,000 different transposons along with their genetic location and nearest genes. The data resulting from this project allowed the generation of specific enhancer traps associated with approximately 40% of the *Drosophila* genes and the potential ability to perform mutagenesis experiments by gene targeting (Bellen et al., 2004).

1.1.3.2 UAS/GAL4 system

Probably the most used and versatile genetic tool in *Drosophila* is the UAS/GAL4 system; the second generation of enhancer trapping (Brand and Perrimon, 1993). This system, which is originally extracted from the yeast *Saccharomyces cerevisiae*, consists of two independent transgenes. On the one hand, the GAL4 gene (that encodes a positive regulator of the galactose-inducible genes in *S. cerevisiae*) is inserted either randomly or specifically (depending in the method utilised for integration) into the genome. This integration allows the GAL4 gene to be expressed under the pattern of any gene of interest, or genomic enhancer. On the other hand, the target transgene is attached in a P-element to the *Drosophila* genome, and it possesses GAL4 binding sites or UAS (Upstream Activation Sequence) regions. Both GAL4 or UAS constructs are inactive on their own (in the case of *Drosophila*, each of them is present in an independent parental strain). However, when both fly strains are crossed, their progeny will express both of them (GAL4 and UAS), inducing the GAL4 activator to recognise the UAS site and therefore enhance the expression of any sequence attached to it (Figure 1.1.1). This downstream sequence can be a gene (to cause an overexpression), an RNAi (RNA interference) (to knock down a gene of interest), or it can be a marker gene, such as GFP. Both fly strains containing the GAL4 and UAS transgenes can be generated and kept independently as separate fly stocks. Moreover, many different GAL4 and UAS lines have been created, and they are available in the stock centres (Dietzl et al., 2007). This significant collection of different modified fly lines actively facilitates the alteration of gene expression under almost any desired pattern.

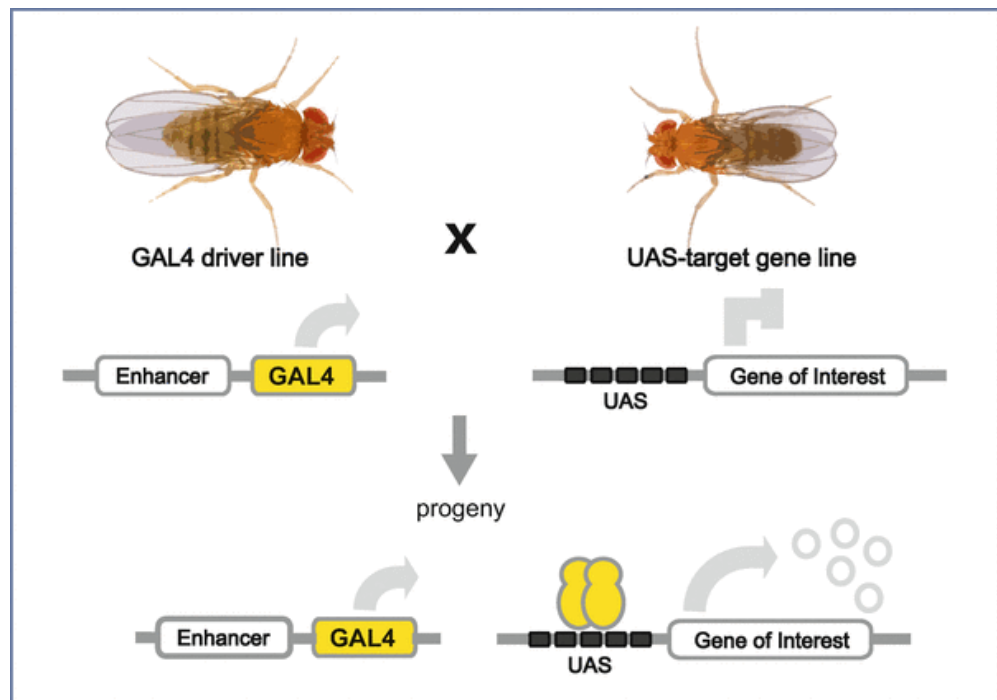


Figure 1.1.1. Schematic diagram of the GAL4 System.

Flies carrying the GAL4 driver transgene (left) are crossed with flies with the UAS-target gene sequence (right). When crossed, only the flies carrying both constructs (GAL4 and UAS-target genes) will result in targeted gene expression. The constructs on their own (present in both parental lines) have no effect on their own in the fly. It is extracted from (Caygill and Brand, 2016).

1.1.3.3 Tubulin GAL80 Thermo-Sensitive system

The GAL4 system can be refined timely and spatially, using the GAL80 Thermo-Sensitive System (TubGAL80^{ts}). The activity of GAL4 can be repressed by interaction with the GAL80 protein, also originated from yeast (Nogi et al., 1984). This protein binds to the C-terminal ends of the GAL4 dimer, making it unable to activate transcription. GAL80^{ts} can be expressed ubiquitously driven by the tubulin-1 α promoter. However, the GAL80^{ts} protein is only active at 18°C (permissive temperature). At this temperature, the dimer GAL80 protein binds strongly to GAL4 protein and a UAS activation sequence, blocking its expression, and therefore preventing the expression of the transgene downstream of the UAS sequence. However, at 29°C (restrictive temperature), GAL80 does not block the function of GAL4, allowing the GAL4 protein to activate GAL4-dependent target gene expression (Figure 1.1.2). Therefore, by combining GAL4 and GAL80 expression patterns, Temporal And Regional Gene Expression Targeting (TARGET, (McGuire et al., 2004)), expression refinement can be drastically improved.

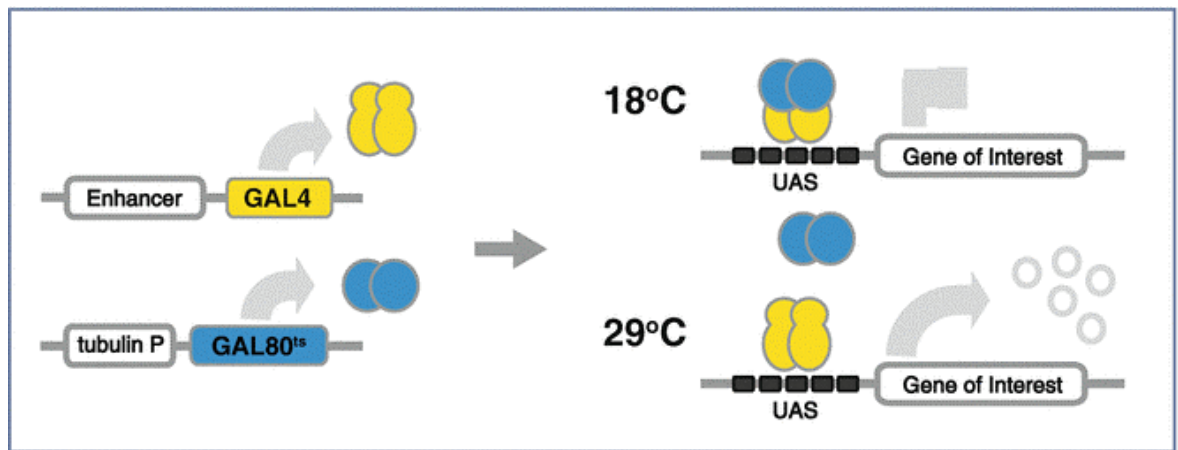


Figure 1.1.2. Scheme of the GAL80^{ts} system.

At 18°C (permissive temperature) GAL80 inhibits GAL4 activity, whereas at 29°C (restrictive temperature) GAL80^{ts} cannot block GAL4, and transgenes can be expressed. Extracted from (Caygill and Brand, 2016).

1.1.4 *Drosophila* life cycle

One of the most significant advantages of *Drosophila* is its short life cycle, of around ten days at 25°C. The length of *Drosophila* development is temperature dependent, which means that at higher temperatures it can develop faster (whole life cycle is ~7d at 29°C) and at lower temperatures, it is slower (~15d at 18°C). Like many invertebrates, *Drosophila* presents a series of developmental phases, starting with the egg stage (around 24h long), three larval stages (Larval stage 1 or L1, Larval stage 2 or L2, and Larval stage 3, or L3), a pupal stage, and the adult stage. The length of these stages (at 25°C) is summarised in Figure 1.1.3. Newly emerged adult females take approximately 10h to be sexually mature, and during this time they present typical characteristics of unmated “virgin” flies, allowing researchers to quickly select them to perform crosses with males of the desired genotype.

Most *Drosophila* adult organs are developed during larval stages and metamorphosis, and at these earlier stages of development, they are called imaginal organs (Beira and Paro, 2016). However, other tissues, including all larval tissues, are developed and functional during the embryonic stage. This is the case of the MTs, which are fully operational by the end of embryogenesis (Beyenbach et al., 2010, Denholm, 2013, Denholm and Skaer, 2009). Even though this organ is completely developed at the end of embryogenesis, recent research has shown that the MTs go through drastic remodeling processes, to assure their correct morphology and survival during metamorphosis and adult stage (Zeng et

al., 2010, Takashima et al., 2013, Singh and Hou, 2009, Singh et al., 2007, Denholm et al., 2013).

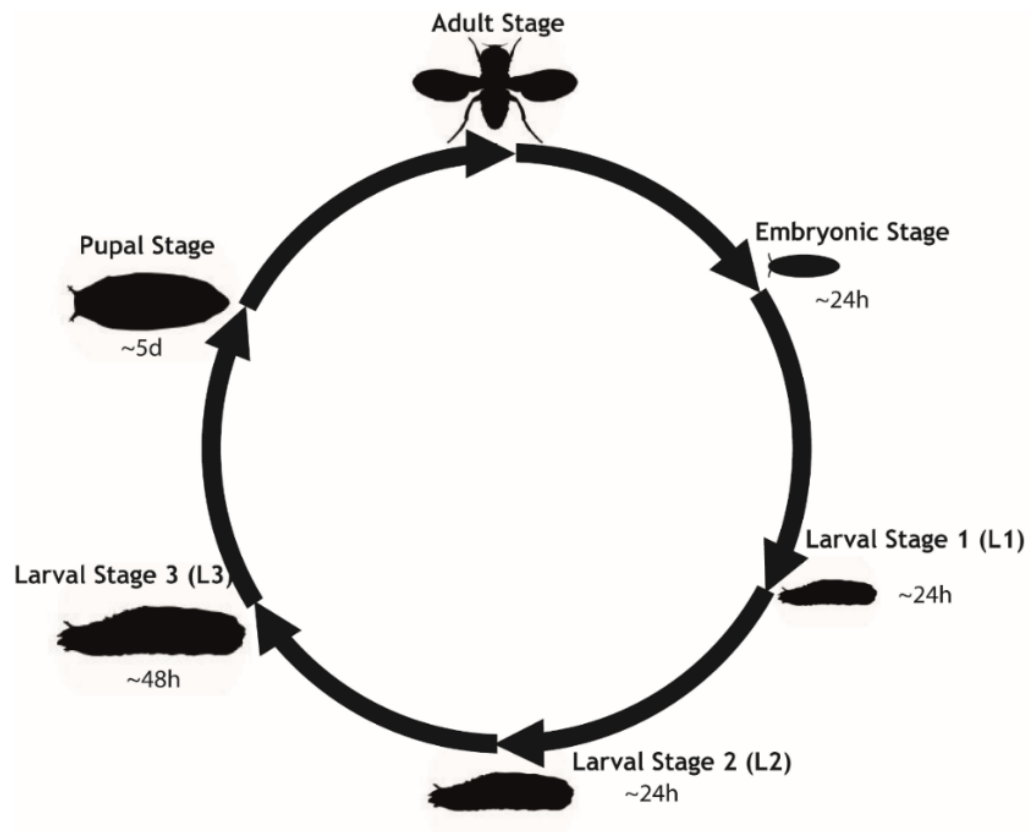


Figure 1.1.3. Stages in the life cycle of *Drosophila melanogaster*.

The entire life cycle of *Drosophila* lasts ~10 days at 25°C. Embryonic development is completed in the first 24h, after which hatching of first instar larvae (L1) occurs. The larvae eat and grow, and moult through three instar larval stages (L1, L2 and L3) before pupariating. During metamorphosis, adult structures are formed, which takes ~5 days at 25°C, and after this stage, and adult fly hatches. Adult and pupal stage pictures are adapted from (Halberg et al., 2016).

1.2 The *Drosophila* Malpighian tubule

1.2.1 Overview

In *Drosophila*, the analogous organs of the vertebrate kidney are the MTs. They are present in numerous types of insects and other terrestrial arthropods, such as myriapods or arachnids (King and Denholm, 2014). The first description of the MTs comes from the 1600s, with the studies and observations of the moth *Bombyx Mori*, performed by the Italian biologist and physician Marcello Malpighi. For his studies, he was considered the father of histology and microscopical anatomy (Bertoloni Meli, 2011). The number, size, and shape of MTs vary significantly between species (e.g. from 2 pairs of tubules in *Drosophila* to more than 200 in the cockroach *Periplaneta Americana* (Wall et al., 1975)). In *Drosophila*, the MTs are single cell layered epithelial tubes, which emerge from

the anterior region of the hindgut (HG), in its boundary with the posterior midgut (PMG), (Denholm and Skaer, 2009). As *Drosophila* has an open circulatory system, MTs float freely through the body cavity allow maximum contact with hemolymph (Dow and Romero, 2010). Additionally, *Drosophila* MTs are not the only organs able to filtrate hemolymph. An additional type of cells, called nephrocytes, have been previously described as the functional and genetical homologue of the human podocytes, present in the glomerulus of the human kidney, and they have been proposed to have roles in the filtration of hemolymph (Weavers et al., 2009).

1.2.2 Structure and function of the MTs

Drosophila MTs produce the primary urine which is generated in the distal region of the tubule and is modified as it passes through more proximal regions until it is excreted from the animal. The MTs also control different ion homeostasis, acid/base, and water balance, and also perform functions of detoxification and immune response (Davies et al., 2012, Denholm, 2013, Dow et al., 1994).

The adult MTs are formed by at least three different types of cells, and each of them has specific functions and physiological characteristics. The most abundant cell type is the Principal Cell (PC, ~120 cells/tubule (Sözen et al., 1997)), which mediates the secretion of Na^+ and K^+ to the tubule lumen, via the vacuolar proton pump V-type H^+ ATPase, located in the apical brush border of the cell (Beyenbach et al., 2010). The other type of cell is the Stellate Cell (SC, ~30 cells/tubule (Sözen et al., 1997)). SCs regulate the Cl^- and water transport (Dow, 2012). They possess ion channels that allow the transport of Cl^- , and aquaporin channels to permit water transport (Kaufmann et al., 2005, Cabrero et al., 2014).

Both PCs and SCs possess receptors for different neuropeptide families, which modulate MT secretion rates and control ion and water homeostasis. Neuroendocrine control of fluid secretion is a complex process, and it is achieved by these two differentiated secretory cell types. Diuretic hormones act through binding of specific receptors in SCs and PCs and stimulating their secretion through specific signalling pathways (Figure 1.2.1). The activation cascades and mechanisms of neuropeptides in the MTs have been extensively

studied and reviewed elsewhere (Cabrero et al., 2014, Terhzaz et al., 2015a, Davies et al., 2014b, Davies et al., 2012, Halberg et al., 2015). A summary of transport functions of both PCs and SCs is provided in Figure 1.2.1, including the different neuropeptides that stimulate each of those cell-types.

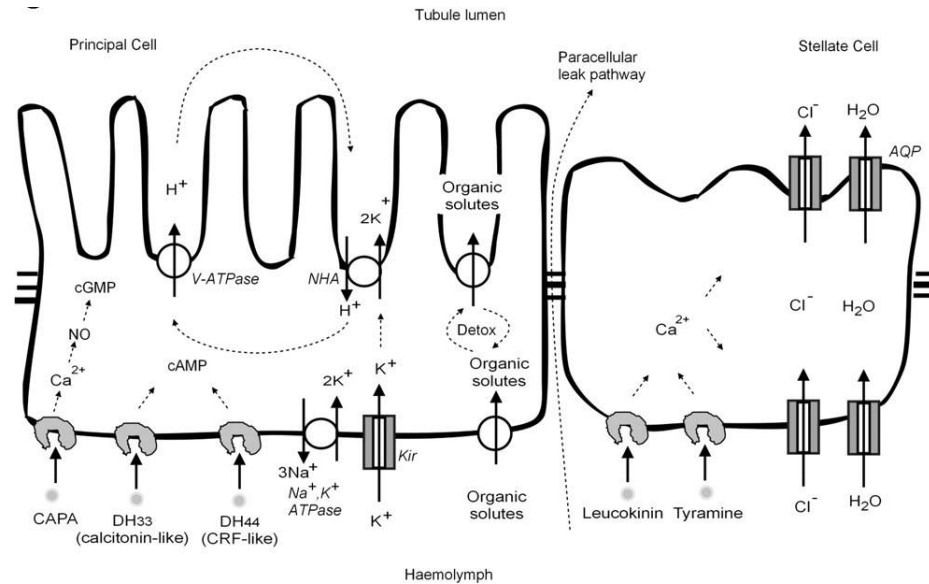


Figure 1.2.1. Schematic overview of the function of a PC (left) and a SC (right) of the main segment of the tubule.

Extracted from (Dow and Romero, 2010).

Furthermore, an additional cell type in the MTs has been recently described, the Renal and Nephritic Stem Cell (RNSC (Singh et al., 2007)). RNSCs migrate from the anterior midgut (AMG) during metamorphosis and are mainly confined in the region of ureters and lower ureters of the adult MTs (Takashima et al., 2013). Previous publications proposed that RNSCs correspond to stem cells in the MTs since they express stem cell markers and they can potentially differentiate into PCs or SCs following injury (Singh and Hou, 2009, Singh et al., 2007, Li et al., 2015). Alternatively, it has been suggested that they could act as endocrine cells because they also express endocrine markers (Sözen et al., 1997).

Adult MTs can also be divided by different domains. They are the ureter, lower tubule, main segment, transitional segment and initial segment (which, in the case of anterior MTs, it is called enlarged initial segment, due to its increased size compared to the posterior initial segment (Figure 1.2.3)). The *Drosophila* MT segments have been genetically characterised by specific gene expression and function. For instance, *Vacuolar H⁺ ATPase* (V-ATPase) expression levels are much lower in the initial and translational segments compared to the main

segment, indicating that this domain is more active regarding cation pumping (Sozen et al., 1997). MTs can also show high asymmetry between posterior and anterior tubules. They display morphological differences, as the anterior pair has an enlarged initial segment, a region capable of transporting and storing more than 90% of the animal's calcium content (Dube et al., 2000). MTs asymmetry can also be studied by observing their differential gene expression. MTs show clear transcriptomic differences regarding their orientation along the body cavity (anterior vs posterior pairs), their cell type or position in the tubule (PCs vs SCs or initial segment vs ureters vs main segment) and their sex (male vs female MTs) (Chintapalli et al., 2012).

1.2.3 Development of the MTs

1.2.3.1 Overview

The development of the MTs occurs principally during embryogenesis (Denholm and Skaer, 2009, Denholm, 2013). However, the tubule goes through different changes during metamorphosis to reach its functional capabilities in the adult stage (Wessing and Eichelberg, 1979). The embryonic development of the MTs can be divided into various stages: (i) the tubule primordia are specified in a subset of cells of the midgut (MG)-HG boundary. (ii) the primordia start budding, followed by a set of controlled cell divisions. (iii) the tip cell is specified, cell proliferation stops, and the tubule primordia start elongating only by cell rearrangement at the same time as a group of cells integrate into the primordia from the caudal visceral mesoderm in a process called mesenchymal-to-epithelial transition. (iv) MTs reach their maturation at the end of embryogenesis, which can be observed by the presence of uric acid crystals in the lumen of MTs of stage 17 embryos and the first stage of the larva (Beyenbach et al., 2010, Denholm, 2013). A summary of the development of the tubule is shown in Figure 1.2.2.

1.2.3.2 Formation of tubule primordia

The development of the MTs starts at stage 10-11 of embryogenesis, where the independent expression of the Transcription Factors (TFs) *Krüppel* (*Kr*) and *cut* (*ct*) is required in a subset of cells of the HG in their boundary with the PMG for the tubule cell specification (Blochlinger et al., 1990, Gaul and Weigel, 1990)

(Figure 1.2.2A). In double mutants *Kr* and *ct*, the MTs fail to differentiate and emerge from the HG (Hatton-Ellis et al., 2007). These two factors act in concert to specify this region, and their pattern of expression is also dependent on other TFs, such as *tailless*, *forkhead* and Wingless signalling (Hatton-Ellis et al., 2007, Ainsworth et al., 2000).

1.2.3.3 Tubule primordia eversion

After the formation of the tubule primordia, the expression of *Kr* refines into four groups of cells, which start to form buds. Of these four everting buds of cells, the ventral pair will become the posterior MTs, and the lateral pair, the anterior MTs. After this stage, the tubule primordia start to evert from the HG. This process is controlled by *decapentaplegic* (*dpp*), one of the main components of TGF- β signalling, which is highly expressed in the anterior tubule primordia compared to the posterior one (Hatton-Ellis et al., 2007). High TGF- β signalling in the anterior tubule primordia induces the expression of *dorsocross*, and expression of *brinker* in the posterior primordia, which will repress signalling. Any disruptions in this differential signalling between both pair of tubule primordia will induce defects in the eversion of the tubules (Hatton-Ellis et al., 2007).

1.2.3.4 Tip cell selection and regulated cell proliferation

When the tubule primordia are everting, a single pair of cells at the tip of each branch arises by a single division of a progenitor cell (Hoch et al., 1994, Skaer, 1989). Although both of them have the same potential to become tip cells, only one adopts this role, and it is identified by competence by lateral inhibition through the Notch pathway and start secreting Epidermal Growth Factor (EGF) (Kerber et al., 1998, Sudarsan et al., 2002). Signalling from this primordium will be received through the EGFR pathway and activation of Spitz and promote cell division (Figure 1.2.2B). After three or four cycles of cell division, it ceases, and the tubule starts a phase of elongation. At this stage, cell number is fixed, with ~140 cells in the anterior tubule pair and ~105 cells in the posterior pair (Ainsworth et al., 2000).

1.2.3.5 Tubule elongation

By stage 12 of embryogenesis, cell division ceases, and the tubule primordia start a robust elongation process (Figure 1.2.2C and D), which will increase their length up to 4 times and their number of cells in the tubule circumference drops from ~10 cells to 2 cells. This achieved through cell intercalation and convergent extension, processes that are regulated by different TFs (*ct*, *trachealess* and *ribbon*), to control the expression of different cytoskeletal and polarity genes (*crumbs (crb)*, *moesin*, *crossveinless-c* or *zipper*), and induce changes in myosin and actin cytoskeleton, required for tubule extension independently of other tissues (Kerman et al., 2008, Hatton-Ellis et al., 2007, Jack and Myette, 1999, Campbell et al., 2009). During this process, the MTs move in a very fixed and stereotypical manner; the posterior pair projects backwards to the HG and will finish the elongation being attached to it. The anterior pair elongates through the abdominal thoracic region and will end up fastening to the alary muscles (Figure 1.2.2D).

This process is still regulated by TGF- β and EGFR signalling (Saxena et al., 2014), the activity of PDGF/VEGF pathway and secretion of type IV collagen, which stimulates BMP signalling pathways (Bunt et al., 2010). In addition, a recent study showed that Wg signalling is also required for the proper proximo-distal orientation of the developing MTs (Beaven and Denholm, 2018).

Tip cells are crucial for this elongation and migration processes through the body cavity. Genetically or physically alterations of these tip cells (inducing a lack of tip cells, or two tip cells instead of one in each tubule) induce problems in the positioning of the tubule, causing failure of fluid homeostasis and premature lethality in the adult stage (Weavers and Skaer, 2013). The kink region of the developing tubule is also vital, and the interaction of this region with Dpp guidance cues from the visceral MG (Bunt et al., 2010, Weavers and Skaer, 2013). Previous literature also suggests an implication of the Extracellular Matrix (ECM) in the development and elongation of the MTs ((Bunt et al., 2010)). Similar to the development of other organs (Jack and Myette, 1999), basal ECM proteins may play a role in the guidance of MT elongation. Additionally, the apical or luminal ECM may play a role in determining the shape and migration of the MTs, in a similar way to other tissues, such as the embryonic trachea (Araújo

et al., 2005). However, it has been previously pointed out, this must be a temporary process, since the MTs are not show the presence of chitin during their development, unlike the trachea or the cuticle (Denholm, 2013).

1.2.3.6 Stellate Cell migration and integration

During elongation, cells are recruited from the caudal visceral mesoderm and migrate to the developing MTs to become SCs, through mesenchymal-to-epithelial transition (Figure 1.2.2C). During SC intercalation, cells of the tubule primordia maintain their epithelial phenotype and polarity, through the activity of *crb*. SC migration is tightly controlled by the TF *hibris* (*hbs*), an Immunoglobulin-like protein, homolog of the human *NEPHRIN*. Mutant flies for *hbs* fail to integrate the SCs into the MTs, resulting in a significant reduction of SCs in adult MTs resulting in premature death (Denholm et al., 2003). In addition, these MTs show a decrease in fluid secretion in response to SC-specific neuropeptides, such as leucokinin (Denholm et al., 2003). Recent studies have demonstrated that Ecdysone signalling is also necessary for the correct integration and allocation of the SCs in the MTs through the EcR-B2 isoform, and blocking Ecdysone signalling in the developing MTs induces strong morphological and physiological defects (Gautam et al., 2015, Gautam and Tapadia, 2010).

1.2.3.7 Tubule maturation

Towards the end of embryogenesis (~stage 17), the MTs are fully developed and functional, shown by the presence of uric acid crystals in the lumen of the MTs (Figure 1.2.2E). However, these crystals will not be cleared until the adult emergence. By the end of embryonic development, MTs already show a pronounced proximo-distal axis (revealed by the presence functional regions differentiated along the tubule length), that is vital for the adult MT functions;

the distal tubule produces primary urine, and it travels and is modified through more proximal areas of the tubule, (Skaer et al., 1990, Beyenbach et al., 2010).

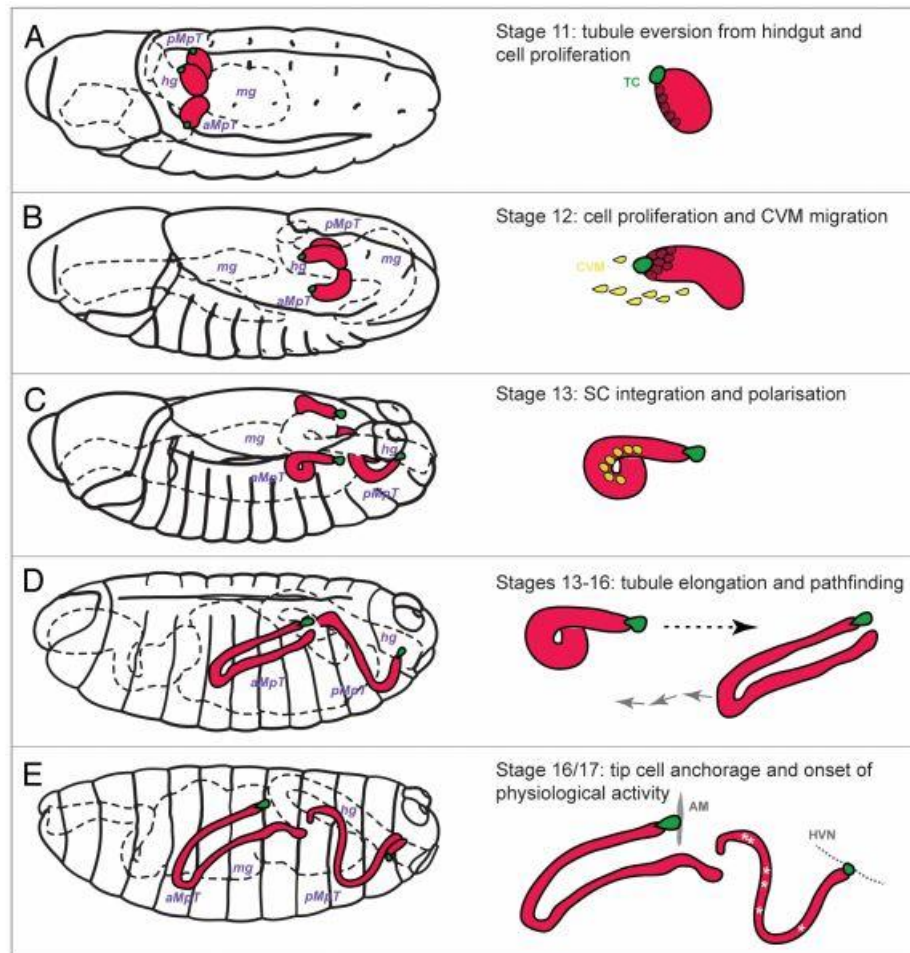


Figure 1.2.2. Scheme of the embryonic development of the MTs of *Drosophila*

(A) At stage 11, MTs emerge from the embryonic hindgut into four buds. (B) At stage 12, tubule primordia go through a synchronised series of cell proliferation, and cells from the caudal visceral mesoderm (yellow) migrate to the tubule. EGF secretion from the tip cell (marked in green) controls cell proliferation at stage 12. (C) At stage 13, cell division ends and SCs integrate into the developing tubule. (D) From stage 13 to 16, the tubule elongates only by cell intercalation. This elongation process is highly stereotypical. (E) By the end of embryogenesis, the tip cells are anchored to its final positions (into the alary muscle, AM in the anterior tubules, and to the hindgut visceral nerve, HVN, in the case of posterior tubules). Physiological maturation can be observed when uric acid crystals are present in the lumen of the posterior pair. Extracted from (Denholm, 2013) with the permission of Dr Barry Denholm.

1.2.3.8 Post-embryonic development

Although the MTs are fully developed by the end of embryogenesis, drastic changes occur at later stages of development, including changes of MT size that are majorly due to cell size increase and endoreduplication of DNA (Denholm, 2013). Interestingly, the MTs survive apoptosis and remodelling events during metamorphosis compared to other tissues of the digestive tract (Wessing and

Eichelberg, 1979). However, the MTs go through different maturation processes during the pupal stage. For instance, pupal SCs change their shape dramatically from a cuboidal to their typical stellate form (or bar-shape in the initial and transitional segments, especially in the anterior tubules). This process is regulated by *tsh*, and downregulation of this gene induces severe reduction in the functional capabilities of SCs, and consequently, lethality (Denholm et al., 2013). Also, the MTs shrink to almost half their size during metamorphosis and form a bulb in their initial segments, and elongate again before hatching. As a consequence, the adult initial segment becomes reduced in size compared with the larval initial segment (Wessing and Eichelberg, 1979).

1.2.3.9 Renal Stem Cell migration and adult tubule maintenance

The third main cell population in the MTs, the RNSCs, integrate to the MTs during the pupal stage. In the adult stage, this cell population is retained only in ureters and lower ureters, and they present different gene expression profiles compared to both PCs and SCs (Singh et al., 2007). For example, RNSCs are negative for both of main segment markers for differentiated tubule cells such as *ct* or *tsh*, but express other markers characteristic of stem cells; they show strong activity of JAK-STAT signalling, express *escargot* (*esg*, a Snail-type TF, involved in stem cell maintenance (Singh et al., 2007, Li et al., 2017)), *hindsight* (*hnt*, another TF involved in stem cell maintenance, (Bohère et al., 2018)), and *armadillo* (*arm*, homolog of β -catenin in *Drosophila* (Brunner et al., 1997)). Moreover, RNSCs express Notch signalling components, such as *delta* (*dl*), and components of EGFR/MAPK signalling, which act independently from JAK-STAT signalling (Li et al., 2015) to regulate their proliferation.

RNSCs originate from a subset of cells called adult MG progenitors that migrate posteriorly from the PMG and colonise the ureter and lower ureter of the adult tubule during metamorphosis. This migration depends on the activity of the GTPase Rac1, and cells with loss of Rac1 fail to migrate to the MTs (Takashima et al., 2013). Several genes have been identified to control the proliferation and survival of RNSCs. The tumour suppressors *salvador* (*sav*) and *scribble* (*scrib*) regulate the proliferation of RNSCs through blocking the activity of Ras signalling. Loss of any of these genes induces tumorous growth and uncontrolled cell proliferation in the MTs (Zeng et al., 2010). In parallel, Shavenbaby (*Svb*)

physically interacts with the TF Yorkie (Yki, a nuclear effector of the Hippo pathway) to ensure the survival of the RNSCs, partially by regulating the expression of the inhibitor of apoptosis *DIAP1* (Bohère et al., 2018).

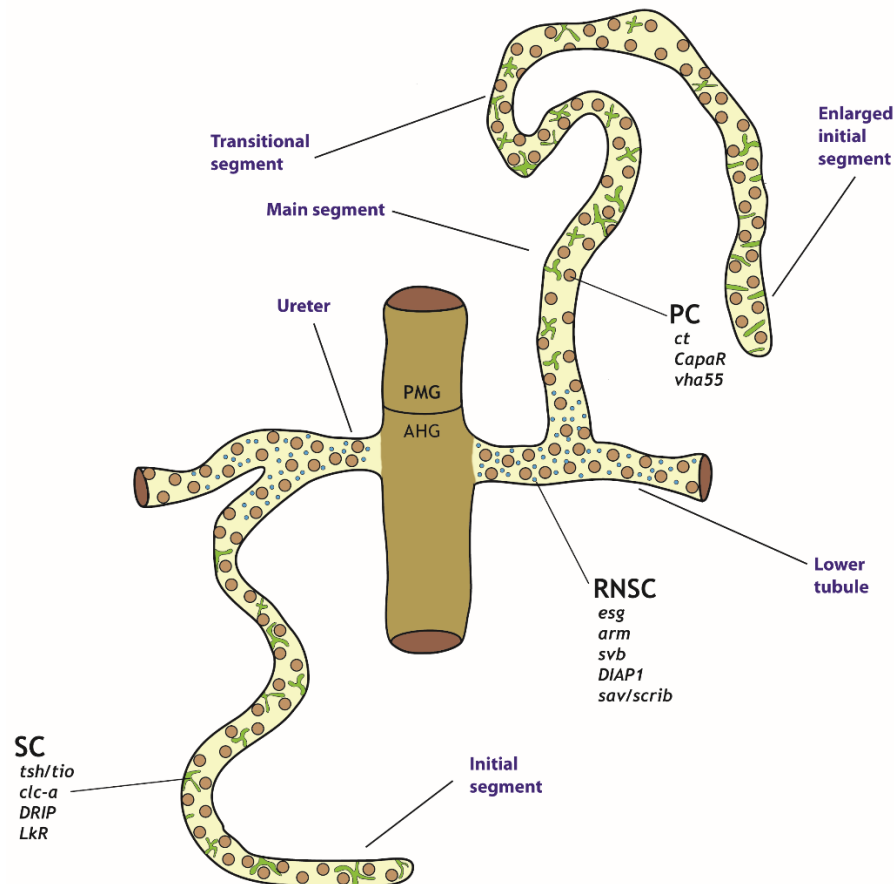


Figure 1.2.3. Schematic diagram of the adult MTs.

Schematic structure of the adult posterior (left) and anterior (right) pairs of MTs. It includes critical gene expression of the three different types of cells. PCs express *ct* (Hatton-Ellis et al., 2007), *CapaR* (Terhzaz et al., 2012), and the V-ATPase *vha55* (Allan et al., 2005). SCs express *tsh* and *tio* (Denholm et al., 2013), *clc-a* and *DRIP* (Cabrero et al., 2014), and *LkR* (Terhzaz et al., 1999). RNSCs, which are limited to ureters and lower ureters, express several stem cell markers, such as *esg*, *arm*, (Singh et al., 2007), *svb* and *DIAP1* (Bohère et al., 2018) and *sav* and *scrib* (Zeng et al., 2010). Only the anterior pair shows the presence of SCs with bar shape, called bar-cells. The different regions are also indicated, in blue. PMG: Posterior Midgut, AHG: Anterior Hindgut.

1.2.4 Transcriptomic profile of the adult Malpighian tubules

Examining the transcriptome profiles of the different organs and tissues of *Drosophila* has been a robust method to investigate the mechanisms underlying the developmental and functional processes in *Drosophila* and other systems by reverse genetics. Transcriptomic studies are especially useful when they are performed in a tissue-specific manner, which is fundamentally the function of the online tool FlyAtlas (Chintapalli et al., 2007). It consists of an extensive organ-specific transcriptomic study in *Drosophila*, allowing the identification of

any pattern of expression of virtually any *Drosophila* gene in both L3 and adult stage (in the adult, it also differentiates males and females) and compares its enrichment in a specific tissue with the whole fly. The first version of FlyAtlas was based on Microarray data. However, the recent release of FlyAtlas2 is based on RNA-sequencing data, which allows transcript differentiation, and the data for somatic tissues is presented as separated in male and female adult flies. Also, FlyAtlas2 includes information of microRNA (miRNA) expression and other RNA genes present in the release 6 of the *Drosophila* reference genome (Leader et al., 2018).

The MTs of *Drosophila* present a particular transcriptomic profile, which has been extensively described (Wang et al., 2004) and can be confirmed using FlyAtlas. Genes encoding transporters are significantly enriched in the MTs, such as ABC, phosphate, or zinc transporters, potassium channels, chloride channels and aquaporins. Second messengers, such as cAMP, cGMP and calcium signalling transducers (such as *Itp-r83A*, or *CapaR*) are enriched as well in the MTs (Figure 1.2.1). Also, cytochrome p450 and glutathione transferase genes are enriched in the MTs, further confirming their functions in detoxification, similar to the human kidney and liver. (Yang et al., 2007). Other genes related to human diseases are strongly upregulated in the MTs, such as *rosy* (Xanthinuria type I) or *Spat* (Hyperoxaluria, type I). Examples of renal diseases discovered in the *Drosophila* MTs include the investigation of mutations of V-ATPase subunits (e.g. proximal renal tubular acidosis) (Allan et al., 2005) or the study of polycystic kidney disease by the study of the evolutionary conserved RNA-binding protein Bicaudal C (BicC) (Gamberi et al., 2017) among other genetic diseases (Wang et al., 2004). Altogether, these investigations provide robustness to future studies that could be performed with these genes using *Drosophila* as a model.

Finally, more than 30 TFs are significantly enriched in the adult MTs (Table 1.2-1). Some of those genes have been extensively studied in this tissue (such as *ct*, *tsh* or *fkh*). However, there are other TFs on this list that have not been investigated in the MTs. Identifying the most critical TFs from this list which could have a role in the development of the MTs, is one of the principal investigations of this PhD thesis.

Symbol (Wang et al., 2004)	Current symbol (2019)	Signal	Enrichment	References in relation to MTs
<i>CG10278</i>	<i>GATAe</i>	175 ± 7	24.1 ± 11.5	-
<i>CG5093</i>	<i>Doc3</i>	50 ± 4	19.3 ± 6.2	(Hatton-Ellis et al., 2007)
<i>pnt</i>	<i>pnt</i>	63 ± 5	17.5 ± 4.8	(Kerber et al., 1998)
<i>CG2779</i>	<i>Muc11A</i>	5771 ± 317	16.8 ± 0.7	-
<i>Ptx1</i>	<i>Ptx1</i>	183 ± 8	12.7 ± 2.2	-
<i>Ets21C</i>	<i>Ets21C</i>	51 ± 17	9.8 ± 3.1	-
<i>CG4548</i>	<i>XNP</i>	91 ± 4	8.8 ± 4.3	-
<i>HLH4C</i>	<i>HLH4C</i>	6 ± 1	7.7 ± 6.9	-
<i>fkh</i>	<i>fkh</i>	266 ± 26	7.2 ± 1.1	(Maruyama et al., 2011, Murakami et al., 1999)
<i>hth</i>	<i>hth</i>	162 ± 13	7.2 ± 0.7	(Zohar-Stoopel et al., 2014)
<i>CG4566</i>	<i>Cnc</i>	17 ± 2	7.1 ± 4.2	-
<i>bowl</i>	<i>bowl</i>	71 ± 5	7.1 ± 0.7	-
<i>CG4037</i>	<i>Not present</i>	5 ± 1	6.7 ± 2.4	-
<i>tap</i>	<i>Tap</i>	5 ± 1	6.0 ± 3.0	-
<i>CG6913</i>	<i>Fer3</i>	5 ± 2	6.0 ± 5.5	-
<i>CG3950</i>	<i>CG34417</i>	287 ± 21	5.4 ± 0.9	-
<i>Awh</i>	<i>Awh</i>	21 ± 4	4.8 ± 1.4	-
<i>CG1162</i>	<i>CG1162</i>	8 ± 1	4.7 ± 2.1	-
<i>ct</i>	<i>ct</i>	145 ± 12	4.6 ± 0.8	(Hatton-Ellis et al., 2007, Ainsworth et al., 2000, Jack and Myette, 1999)
<i>CG14202</i>	<i>CG32532</i>	10 ± 1	4.6 ± 1.5	-
<i>tsh (ae)</i>	<i>Tsh</i>	65 ± 5	4.6 ± 0.8	(Denholm et al., 2013, Denholm et al., 2003)
<i>CG9952</i>	<i>Ppa</i>	45 ± 11	4.5 ± 0.6	-
<i>sv</i>	<i>sv</i>	16 ± 2	4.3 ± 1.8	-
<i>fd59A</i>	<i>fd59A</i>	11 ± 3	4.3 ± 1.7	-
<i>CG11914</i>	<i>lmpt</i>	31 ± 4	4.2 ± 1.7	-
<i>slp2</i>	<i>slp2</i>	4 ± 2	4.1 ± 3.1	-
<i>Lim3</i>	<i>Lim3</i>	13 ± 3	4.0 ± 1.1	-
<i>CG6419</i>	<i>Sox21b</i>	18 ± 3	4.0 ± 0.4	-
<i>Tis11</i>	<i>Tis11</i>	337 ± 17	3.9 ± 0.6	-
<i>nvy</i>	<i>nvy</i>	27 ± 4	3.9 ± 1.1	-

Table 1.2-1. Transcription factors enriched in the adult MTs.

TFs are ranked by their enrichment, which results from the comparison with their tubule-specific with whole-fly expression levels. Data extracted from (Wang et al., 2004). The current gene symbols are also included, according to the *Drosophila melanogaster* genome release 6.27 (2019).

1.3 Transcription Factors

1.3.1 Overview

TFs are proteins that can modify the expression of downstream genes by binding to specific sequences in the DNA, usually at the 5' end of those genes. The primary function of TFs is to ensure the proper spatiotemporal regulation of downstream genes (Latchman, 1997, Thomas and Chiang, 2006). TFs can influence gene expression either positively or negatively, and there are several

families of TFs, which can be classified by the structure of their DNA binding domains (DBD). The study of the TF function and specificity can be performed using *Drosophila* as a compelling model, due to its genetic versatility and similarity of its genome with the human genome. The different families of TFs can be classified by the structure of their DBD domains. There are more than 15 types of DBDs and the most studied TF types according to the structure of their DBDs have been reviewed in previous studies. (Bürglin and Affolter, 2016, Takatsuji, 1998, Lentjes et al., 2016, Sharrocks, 2001, Xie et al., 2016, Vinson et al., 1989, Toledo-Ortiz et al., 2003).

Interestingly, recent research has elucidated other ways to alter gene expression with ‘synthetic’ TFs, such as Transcription activator-like effector nucleases, or TALENs (Maeder et al., 2013, Perez-Pinera et al., 2013b) or CRISPR-cas9 based TFs (Perez-Pinera et al., 2013a).

1.3.2 Transcription factors, development and disease

TFs are present in all living organisms, from bacteria to humans. In metazoans, TFs are involved in several developmental processes, such as the determination of the body pattern (Bejsovec and Wieschaus, 1993), or the development of embryonic and imaginal structures (Kuo et al., 1997, Lentjes et al., 2016, García-Bellido, 2009). Loss or altered expression of any of these TFs induces severe defects in the development, and, in most cases, embryonic lethality. For these reasons, understanding the mechanisms which dictate TF specific binding to DNA is vital to uncover the molecular and genetic processes of development and disease (Lee and Young, 2013). These mechanisms should not be limited to study protein interactions with the DNA. Previous research has identified novel functions of different non-coding RNAs (such as microRNAs or piwi-interacting RNAs) in the regulation of gene expression, conferring protection to multiple diseases (Care et al., 2007, Ebert and Sharp, 2012, Aravin et al., 2007).

1.3.3 Conservation of transcription factor binding specificities across evolution

In the last years, the study of TFs has increased its significance in model organisms such as *Drosophila*, as high-throughput technologies and methods identified that despite several hundreds of millions of years of evolution, most of

the DNA motifs found in humans and mice are also recognised by *Drosophila* TFs. This surprising discovery indicates that TF evolution and diversification is a process that can run relatively slowly, and it can be clearly and easily studied in simpler organisms like *Drosophila*. Therefore, flies have been used as a bed-test for further studies investigating the function of different TFs in mice or humans (Merika and Orkin, 1993, Nitta et al., 2015).

1.3.4 Transcription factors enriched in the *Drosophila* MTs

Several families of TFs are enriched in the *Drosophila* MTs (Table 1.2-1) (Wang et al., 2004)). Almost half of the TFs in this list have already been characterised in the MTs, including *ct* and *tsh*, two of the most studied in the embryonic development of this tissue (Denholm et al., 2003, Ainsworth et al., 2000). The expression of some of these genes in the MTs has been confirmed by using enhancer trap reporters that highlight their expression, only a few of them were expressed in all regions of the MTs. By contrast, most of these TFs seem to be expressed in specific areas or groups of cells (Wang et al., 2004, Denholm et al., 2003).

However, the other half of TF genes described in this table have not yet been studied for their putative function in the MTs, and some of them have not been yet investigated (as they are still known as their CG numbers). However, several of their functions have been discovered since the date of which Table 1.2-1 data was published. Maybe the clearest example is the first one on the list: *CG10278*, or *GATAe*, a GATA factor that is crucial for the development of the embryonic MG, and the maintenance of this tissue in the adult stage. Loss of *GATAe* in embryos causes the MG (which is an endodermal tissue) to go through an ectodermal fate (Okumura et al., 2005). *GATAe* is also required in the different cell types of the adult MG (Dutta et al., 2015, Zhai et al., 2017, Okumura et al., 2016, Zhai et al., 2018, Dobson et al., 2018). For these reasons, the enrichment of *GATAe* in the MTs is intriguing, as this tissue is ectodermal in origin.

1.4 The GATA family

1.4.1 Gene family and structure

The GATA family of TFs is present in animals, plants and fungi (Vonk and Ohm, 2018) and it is involved in a diverse range of processes in all of them (Lentjes et al., 2016, Patient and McGhee, 2002). Although they have different functions, the principal feature of the TFs of the GATA family is that they possess one (predominantly in invertebrates) or two (predominantly in vertebrates) zinc finger DBD, with a consensus sequence of Cys-X₂-C-X₁₇-Cys-X₂-Cys followed by a basic region (Lentjes et al., 2016). Phylogeny studies suggest that the second Zinc Finger (C-terminal) is present in both vertebrates and invertebrates, indicating that one Zinc Finger could have emerged by a genetic duplication from the other (He et al., 2007). However, during evolution, GATA factors seem to have gone through high divergence regarding their structure and family gene number across phyla. Following these observations, it is remarkable to observe that there are 6 GATA factors in vertebrates, 11 in *C. elegans* and 5 in *Drosophila* (Lentjes et al., 2016, Patient and McGhee, 2002). Despite this difference in gene number between *Drosophila* and vertebrates, general functions of GATA factors are similar in both vertebrates and invertebrates. Several exact potential orthologs between invertebrates and vertebrates have been identified, but it is still difficult to assign all the functions of an invertebrate GATA factor to a unique vertebrate GATA factor. For instance, *serpent* (*srp*) and *GATAe* in flies have been proposed to be homologues to the *GATA4*, 5 and 6 group in humans (see 1.4.2), as both groups are involved in the development of the endoderm. Nevertheless, *srp* is also crucial in haematopoiesis, a process that is controlled by *GATA1*, 2 and 3 in humans (Gillis et al., 2008).

Furthermore, in both vertebrates and invertebrates, GATA factors can also interact with each other, and different members of the family can be involved in gene compensation, i.e. the loss of a GATA gene can be compensated by another GATA gene to perform some of their functions. For example, *GATA1* can rescue early functions in erythropoiesis in *GATA2* mutant *Xenopus* (Mead et al., 2001), which indicates overlapping in their role in erythropoiesis. Another example is *srp* and *GATAe* which, in *Drosophila*, are required for the specification and

differentiation of the of the embryonic MG. It has been shown that *srp* expression is required for the terminal differentiation of the MG, by activating the expression of *GATAe* to ensure the expression of endoderm-specific downstream genes (Okumura et al., 2005).

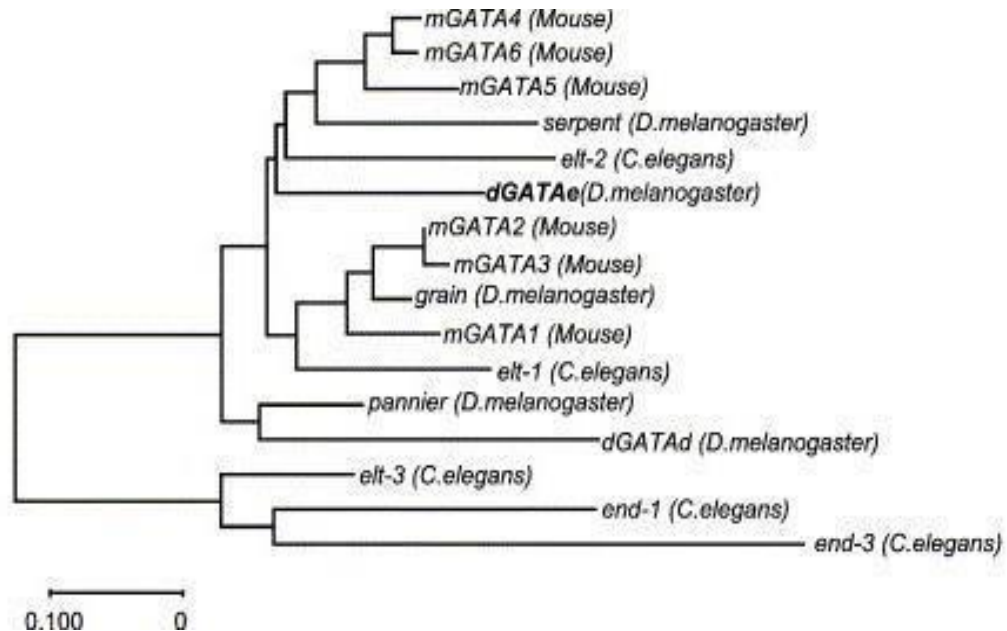


Figure 1.4.1. Cladogram of the different GATA factors in vertebrates and invertebrates. This cladogram includes data from *C. elegans*, *Drosophila*, and mouse, but it excludes human GATA factors. *Drosophila* GATAe (**dGATAe**) is shown in bold. Extracted from (Okumura et al., 2005).

1.4.2 GATA factors in vertebrates

Homo sapiens possesses 6 GATA factors, which can be subdivided into major groups: GATA factors 1, 2 and 3 are necessary for the development of ectodermal and mesodermal tissues, such as the nervous or the hematopoietic system. On the other hand, GATA factors 4, 5 and 6 are highly involved in the growth and the differentiation of mesodermal and endodermal tissues, and the differentiation of embryonic stem cells, adult epithelia or the cardiovascular system (Lentjes et al., 2016, Pihlajoki et al., 2016). Also, they have been associated with diverse types of cancer (Zheng and Blobel, 2010).

1.4.2.1 Developmental functions

GATA 1, 2 and 3

This subgroup of GATA factors has a critical role in the development of blood lineages since mutations of any of these TFs in mice are embryonic lethal due to hematopoietic deficiencies.

GATA1 is specifically involved in the production of the erythroid precursor cell population, and crucial for the maintenance of adult hematopoietic stem cells (Shimizu et al., 2001, Lentjes et al., 2016). *GATA1* deficiencies in humans induce hematopoietic abnormalities, such as anaemia or leukaemia. *GATA1* also represses *GATA2* in normal erythropoiesis. However, in *GATA1* mutants, *GATA2* can partially rescue hematopoietic problems (Ohneda and Yamamoto, 2002, Weiss et al., 1994). In addition, *GATA1* is involved in testis development, but null mutations of *GATA1* do not induce a phenotype in the adult testis, possibly indicating gene compensation by other factors (Lindeboom et al., 2003). *GATA2* is expressed in hematopoietic progenitor cells, and it is required in erythropoiesis. Mutant mice for this gene present a severe reduction in blood cells during embryogenesis. Furthermore, *GATA2* deficiencies in humans are related to blood diseases, such as acute myeloid leukaemia (Rodrigues et al., 2008, Tsai et al., 1994). *GATA2* is also expressed in the developing nervous system in mice, in different populations of neurons. Other studies also have shown a possible interaction of *GATA2* and *GATA3* in the development of the neural tube (Karunaratne et al., 2002).

GATA3 is expressed in hematopoietic stem cells and T-lymphocytes during development. It regulates hematopoietic stem cell production and differentiation, and deficiencies of *GATA3* induce low numbers of this type of cells (Hosoya et al., 2010). Thus, *GATA 1, 2 and 3* present a similar pattern of expression in the developing hematopoietic system. Furthermore, ~10% of breast cancers present mutations in *GATA3*, indicating a role of this gene in the mammary gland, as a tumour suppressor. In addition, *GATA3* has crucial role in vertebrate kidney development as reduction of *GATA3* in humans is a cause of hypoparathyroidism, sensorineural deafness and renal disease (HDR) syndrome (Ali et al., 2007, Ferraris et al., 2009), but is also expressed and required in the nephric duct during the development of the human kidney (Chia et al., 2011, Grote et al., 2006). In the adult kidney, *GATA3* has also been associated with renal cell carcinoma. Recent studies demonstrated that high methylation of CpG

islands of *GATA3* was associated with advanced levels of metastasis in this cancer (Peters et al., 2014b).

GATA4, 5 and 6

GATA4 is necessary for appropriate sex differentiation in mammals, for the development of the lungs, and acts as a negative regulator for astrocyte differentiation in the adult central nervous system (Agnihotri et al., 2009, Ackerman et al., 2007, Lourenço et al., 2011). It is also required in the intestinal system, needed for the appropriate differentiation of mouse embryonic stem cells in the formation of the yolk sac, and to maintain the identity of jejunum and ileum (Soudais et al., 1995). In addition, *GATA4* is required to protect cardiomyocytes from drug-induced apoptosis, and directly targets the human apoptosis inducer *Bcl2* (Kobayashi et al., 2006, Kobayashi et al., 2010, Aries et al., 2004).

GATA5 is expressed in the developing heart, and each of the specific isoforms of this gene is involved in a different part in the development of this organ (Laforest et al., 2011, Jiang et al., 2013). Null mutations for both isoforms of *GATA5* induce hypoplastic heart disease. Heterozygous mutations of *GATA5* have been related to familial atrial fibrillation disease in humans, as well as congenital heart disease. It has also been reported as a potential biomarker for colorectal cancer (Akiyama et al., 2003, Hellebrekers et al., 2009). *GATA5* is expressed during urogenital development, and females with loss of *GATA5* present strong defects in their urethra and vagina (Molkentin et al., 2000).

As well as *GATA5*, *GATA6* is required in the patterning of the developing heart and identified as a target of *Wnt2*. Mutations in *GATA6* have been detected in congenital heart disease in mice and humans (Kodo et al., 2009, Morrissey et al., 1996). *GATA6* is expressed throughout the intestinal tract and (along with *GATA4*) is required for the homeostasis of the small intestine in mice (Tian et al., 2010, Morrissey et al., 1996, Beuling et al., 2011). *GATA4* and 6 are expressed in the developing liver and required for the expression of *Hepatocyte Nuclear Factor 4 (HNF4)*, a gene required for the differentiation of visceral mesoderm and liver (Morrissey et al., 1998). Further, *GATA6* is overexpressed in pancreatic cancer, making it an important marker for the detection of this disease (Chia et al., 2014). *GATA6* is also expressed in the urogenital tract but does not have an overlapping function with *GATA5*. However, *GATA4* and 6 are both expressed in a similar pattern in testis and ovaries, and *GATA6* has been

reported as a marker for ovarian pancreatobiliary and colon cancer (Heikinheimo et al., 1997, Belaguli et al., 2010, Kwei et al., 2008). Other studies showed, in contrast, that *GATA6* expression is reduced in colon carcinomas (Haveri et al., 2008).

In conclusion, vertebrate GATA factors are crucial for the development and maintenance of most organs and, at least two of them (*GATA3* and *5*) are expressed in the embryonic and adult kidney. Hence, loss of any of these two genes can cause defects in the development of the kidney and tumorous growth in humans. Potential homologous functions between the vertebrate factors *GATA3/5* and the *Drosophila* *GATAe* could be investigated, as in flies, *GATAe* is the only GATA factor enriched in the developing and adult MTs.

1.4.2.2 Vertebrate GATA factors and cancer

Aside from developmental functions, all GATA factors have significant roles in diverse human cancers, as they can act both as oncogenes, or as tumour suppressors (Zheng and Blobel, 2010).

Mutations in the hematopoietic gene *GATA1* have been associated with megakaryoblastic leukaemia (Li et al., 2005, Wechsler et al., 2002). Also, *GATA2* has been associated with the aggressiveness of prostate cancer (Rodriguez-Bravo et al., 2017, Wu et al., 2014, Vidal et al., 2015). In addition, epigenetic silencing of *GATA4* and *GATA5*, and consequent downregulation of their tumour suppressor target genes is commonly found in colorectal and gastric cancers (Akiyama et al., 2003) and proposed to act as tumour suppressors in *in vitro* models of these cancers (Hellebrekers et al., 2009). Interestingly, GATA factors also play essential roles in human kidney cancers. Both *GATA3* and *GATA 5* are hypermethylated in human renal clear cell carcinomas, (Peters et al., 2014b) and low levels of *GATA5* mRNA have been associated with this type of renal cancer (Peters et al., 2014a), suggesting an indirect tumour suppressor role of these two GATA factors.

Strikingly, human GATA factors can act both as oncogenes or as tumour suppressor genes, depending on the context and the type of cancer. For example, *GATA3* it is a recognised indicator of breast cancer, and it has been

shown that expression of GATA3 is sufficient to stop tumour dissemination in this type of cancer (Kouros-Mehr et al., 2008). However, GATA3 can also induce carcinogenesis in lymphoid precursor cells, and converts double-positive thymocytes into a premalignant state (Nawijn et al., 2001, van Hamburg et al., 2008). GATA6 can also perform opposite functions in tumorigenesis. While it is a recognised tumour suppressor of astrocytoma, (the most common type of primary brain tumour) (Kamnasaran et al., 2007, Whissell et al., 2014), it has been also proposed to act as a potential tumorigenic factor in pancreatic (Kwei et al., 2008) and colon (Shureiqi et al., 2007) carcinomas.

1.4.3 GATA factors in *Drosophila*

There are 5 GATA factors in *Drosophila*: *GATAa*, or *pannier* (*pnr*), *GATAb* or *srp*, *GATAc* or *grain* (*grn*), *GATAd* and *GATAe*. Previous studies in evolution and phylogeny of GATA factors suggest that *srp*, *GATAe* and *pnr* emerged from two independent events of duplication of the single ancestral bilaterian GATA factor from the group 4, 5 and 6 (*GATA4,5*, and 6), as these three *Drosophila* factors are clustered in tandem in a short genomic region of 70kb (Gillis et al., 2008, Okumura et al., 2005).

1.4.3.1 *pannier*

pnr functions have been extensively studied in *Drosophila*. It was first discovered for its involvement in the bristle patterning, regulating the expression of *achaete* and *scute*, and interacting with *u-shaped* (Haenlin et al., 1997). *pnr* is essential for the development of the heart and shown to interact with *Tinman* (Lovato et al., 2015). Other studies have shown that *pnr* is required for the development of the ectoderm (Raftery and Sutherland, 2003).

1.4.3.2 *serpent*

srp has been extensively studied in the developing MG of *Drosophila*. It is required for the endodermal fate of this tissue as *srp* mutant embryos fail to differentiate the terminal regions of the blastoderm into the endoderm and differentiate into ectodermal hindgut instead (Murakami et al., 2005). *Srp* is expressed in early stages in the embryonic MG until stage 11, activating the expression of *GATAe*, which will be required for the terminal differentiation of

this tissue (Okumura et al., 2005). *Srp* also interacts with *u-shaped* to determine blood cell formation (Waltzer et al., 2002), hematopoiesis (Fossett et al., 2003, Spahn et al., 2014) and it is sufficient to promote fat cell differentiation, (Hayes et al., 2001, Sam et al., 1996). Finally, recent studies show that *srp* participates in the control of the circadian clocks together with other TFs such as *clock* and *cycle* (Meireles-Filho et al., 2014).

1.4.3.3 *grn*

grn is involved in several processes in fly development. It is necessary for the formation of the posterior spiracles and adult legs by controlling cell rearrangement (Brown and Hombria, 2000). *grn* is necessary for the axon guidance of the dorsal motoneurons, by cooperating with *even skipped* and regulating *unc-5* (Zarin et al., 2012), and induces the expression of *fasciclin 2* and *neuroglian*, which encode for cell adhesion proteins to guide these motoneurons (Zarin et al., 2014). It has also been proposed that *grn* and *GATAe* together control immune gene expression in the MG, as their patterns of expression overlap in these regions (Senger et al., 2006). Recent studies showed that *grn* expression in the developing embryonic MG is regulated by *srp* (de Madrid and Casanova, 2018).

1.4.3.4 *GATAd*

GATAd is possibly the least studied of all GATA factors in *Drosophila*. According to the BDGP in situ database, *GATAd* mRNA is expressed maternally and then ubiquitously in the embryo, and FlyAtlas data suggests that it is enriched in the brain (in both larval and adult stage), in the thoracicoabdominal ganglion, and the ovary in the adult stage. This pattern of expression suggests that *GATAd* could have a function in the development or physiology of these tissues, but this remains to be explored.

1.4.4 *GATAe*

GATAe has been extensively studied in the last decade for its functions in the MG. Investigation of possible novel roles of *GATAe* in the MTs has been one of the core aims of this PhD project for two main reasons: (i) it is highly enriched in the MTs in all developmental and adult stages (Okumura et al., 2005) (Figure

1.4.2C and Figure 1.4.3). (ii), although *GATAe* has been extensively studied in other tissues in *Drosophila* (see below), little is known about its functions in the MTs despite its abundant expression in this tissue. In addition, its possible homologies with the vertebrate *GATA3* and 5 TFs, involved in the development and maintenance of the human kidney make *GATAe* an exciting gene to study in the fly kidney (see 1.4.2).

1.4.4.1 Structure and expression pattern

The function of *GATAe* was first characterised in *Drosophila* by researchers of Takashi Okumura's Laboratory, in a search for novel GATA factor genes in the *Drosophila* genome. They successfully cloned two genes that contained GATA factor motifs, *GATAd* and *GATAe*. *GATAe* is genetically located in the third chromosome at 89A13-B4, and it is clustered in a genomic region together with *pnr* and *srp* (Okumura et al., 2005, Gillis et al., 2008).

GATAe encodes three different transcripts: the transcript A (the most enriched in the MTs) is 3086 bp, which translates for a protein of 746 amino acids with two typical GATA zinc finger DBDs, according to (Okumura et al., 2005). One of them is in the C-terminal domain, and it is a classical GATA DBD (C-X₂-C-X₁₇-C-X₂-C). The other zinc finger, located in the N-terminal domain (C-X₂-C-X₁₂-C-X₂-C), is shorter. Both zinc finger domains are followed by flanking basic regions, that constitute the whole DBD (Okumura et al., 2005). *GATAe* also encodes for two other transcripts, B and C, composed of 2983 and 2879 bp respectively, and both translate for an identical protein of 731 aa (Figure 1.4.2A).

GATAe has been typically described as an endodermal-specific gene (Okumura et al., 2005, Okumura et al., 2016). Its closest homologues in other animals are *elt-2* in *C. elegans* and *GATA 4, 5 and 6* in mice/humans (Figure 1.4.1). In *Drosophila* embryos, *GATAe* is mostly enriched in the developing MG, where it is vital for this tissue's identity and development, but it is also strongly expressed in the embryonic MTs, a tissue of ectodermal origin (Okumura et al., 2005).

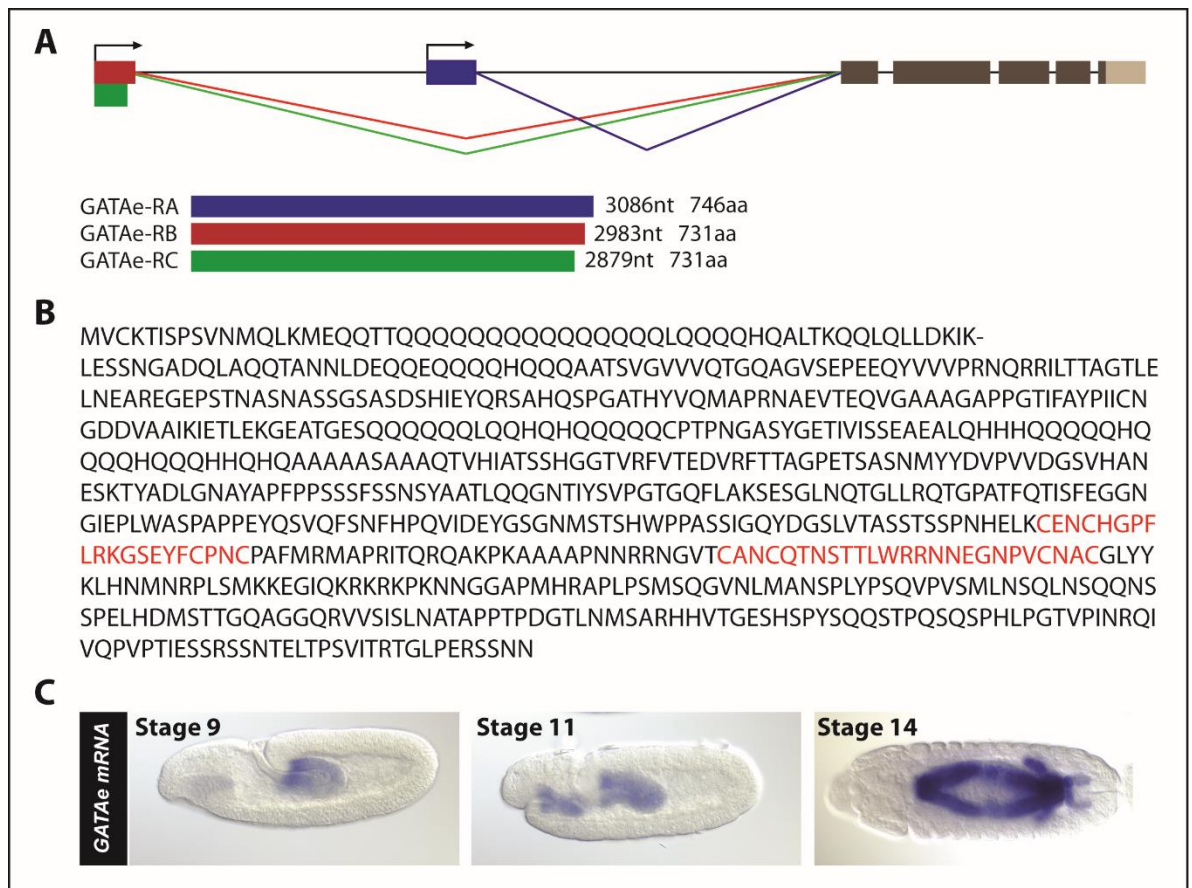


Figure 1.4.2. *GATAe* genomic location and embryonic pattern of expression.

All three *GATAe* isoforms are composed of six exons. The isoforms B (red) and C (green) are identical, with 731 amino acids, whereas the isoform A (blue) is slightly bigger, with 746 amino acids. (B) The predicted amino acid sequence of the isoform A of *GATAe* with zinc finger domains is marked in red, adapted from (Okumura et al., 2005). (C) In situ hybridisation of *GATAe* mRNA in the embryo at three different stages. In early stages (9-11) *GATAe* expression is limited to the MG, and MT primordia, whereas in later stages (15) *GATAe* is expressed across the MG, the proventriculus and the MTs (preparations in C have been extracted from BDGP database).

Interestingly, *GATAe* maintains a similar pattern of expression throughout all developmental stages, being enriched in the MG and MTs of L3 and adult flies, with some expression in the adult male HG (Figure 1.4.3A). Among all three transcripts of *GATAe*, only the transcript A is strongly expressed in the larval and adult MG and MTs, with low expression in the adult hindgut. Also, transcript B is expressed at weak levels in the larval and adult MG and MTs. Strangely, transcript C does not seem to be enriched in any *Drosophila* organ (Figure 1.4.3B). Studies on *GATAe* have mainly been focused on its implication in the developing MG (Murakami et al., 2005, Okumura et al., 2005, Okumura et al., 2007), maintenance of adult MG and intestinal stem cells (ISCs) (Zhai et al., 2017, Okumura et al., 2016, Takashima et al., 2013, Dutta et al., 2015), as a mediator for dietary restriction-induced mediator of lifespan (Dobson et al., 2018) and its role in the immune response of the larval MG (Senger et al., 2006).

A

Tissue	Adult Male		Adult Female		Larval	
	FPKM	Enrichment	FPKM	Enrichment	FPKM	Enrichment
Head	0.0	0.0	0.0	N.A.		
Eye	0.0	0.0	0.0	N.A.		
Brain / CNS	0.0	0.0	0.0	N.A.	0.0	0.0
Thoracoabdominal ganglion	0.0	0.0	0.3	N.A.		
Crop	0.1	0.1	0.1	N.A.		
Midgut	25	11	20	10	15	4.4
Hindgut	2.3	1.0	1.8	N.A.	0.3	0.1
Malpighian Tubules	17	7.7	16	7.9	7.0	2.1
Fat body	pending	—	pending	—	0.2	0.0
Salivary gland	0.4	0.2	0.1	N.A.	0.0	0.0
Heart	pending	—	pending	—		
Trachea					0.1	0.0
Ovary			0.0	N.A.		
Virgin Spermatheca			pending	—		
Mated Spermatheca			pending	—		
Testis	0.0	0.0				
Accessory glands	0.1	0.0				
Carcass	0.5	0.2	0.7	N.A.	0.0	0.0
Rectal pad	0.0	0.0	0.1	N.A.		

B

Transcript		Male															Female															Larval														
Name	ID	Hd	Ey	Br	Tg	Cr	Mg	Hg	Tu	Fb	Sg	Ht	Ts	Ag	Cs	Rp	Hd	Ey	Br	Tg	Cr	Mg	Hg	Tu	Fb	Sg	Ht	Ov	Vs	Ms	Cs	Rp	Ns	Mg	Hg	Tu	Fb	Sg	Tr	Cs						
RA	FBtr0083218																																													
RB	FBtr0303260																																													
RC	FBtr0303261																																													

Figure 1.4.3. FlyAtlas2 reveals *GATAe* expression in L3 and adult stage.

GATAe is enriched in the MG and MTs (A). It is enriched up to 7.7 and 7.9 times in adult male and female MTs, respectively, and also enriched 2.1 times in larval MTs. It shows an 11-fold, 10-fold and 4.4-fold in the adult male, adult female and larval MGs, respectively. Differential expression of the isoforms can also be observed in FlyAtlas 2 (B), and the isoform A is more enriched than the isoform B on C in both the MG and the MTs (Tu). Data extracted from <http://flyatlas.gla.ac.uk/FlyAtlas2/index.html?search=gene&gene=GATAe&idtype=symbol#mobileTargetG>.

1.4.4.2 *GATAe* in embryonic development

In the embryonic MG, *GATAe* is required for the expression of endodermal terminal differentiation genes. Another GATA factor, *srp*, is expressed in the endoderm from stage 7-8 and ceases its expression at stage 11 of embryogenesis (Rehorn et al., 1996). *srp* activates the expression of *GATAe* in the developing MG from stage 7. *GATAe* is necessary to induce downstream gene expression to ensure correct development of the MG. *GATAe* loss in this period will cause that the MG acquires an ectodermal identity by expression of *brachyenteron*, an ectodermal gene that in normal conditions, it repressed by *GATAe* (Okumura et al., 2005, Murakami et al., 2005). In addition, recent publications demonstrated that *GATAe* expression in the embryonic MG does not depend on *grn* (de Madrid and Casanova, 2018).

Furthermore, *GATAe* is expressed in the embryonic MT primordia from stage 10, independently from *srp*. In this tissue, *GATAe* expression is regulated by *Kr* (Okumura et al., 2005). However, the possible functions of *GATAe* in the embryonic MTs have not been yet uncovered.

1.4.4.3 *GATAe* in the larval and adult stages

GATAe has also been shown to be involved in the production of the immune response in the larval MG, as microarray experiments demonstrated that this gene regulates *dipteracin* and *metchnikowin* (two antimicrobial peptide genes). In addition, they suggested that *GATAe* induces an immune response in a Toll signaling-independent manner (Senger et al., 2006).

GATAe is present in all the regions and cell types of the adult MG (Dutta et al., 2015, Okumura et al., 2016). It is crucial for the maintenance and proliferation of ISCs as *GATAe* loss in this cell population resulted in a reduction of epithelial renewal in the adult MG (Okumura et al., 2016, Dutta et al., 2015). Also, it is necessary for the differentiation of enterocytes and enteroendocrine cells, acting both dependently and independently from Notch signalling. Other publications have also shown that *GATAe* actively controls the differentiation of enteroblasts into enterocytes, acting downstream of JAK/STAT signalling and *Sox21a* (Zhai et al., 2017). Furthermore, recent data showed that *GATAe* is required to repress the epithelial shedding of enterocytes in the adult MG upon bacterial infection (Zhai et al., 2018).

As showed in Figure 1.4.3 and Figure 1.4.2, *GATAe* is strongly upregulated in the MTs in all development stages of *Drosophila*. However, whether it has a function in the development of this tissue or not is yet to be studied.

1.5 Aims of the project

Given the importance of TFs in the development and physiology of an organ, it is crucial to identify and study any possible novel roles that they could perform, especially in relation to human development and disease. The *Drosophila* MTs proved to be an excellent model to select and investigate possible phenotypes associated with loss of TF expression.

There are numerous TFs highly enriched in this tissue, and the potential novel functions of many of them have not been yet investigated. On the basis of these findings, four different major points of this PhD thesis have been proposed. They were:

- To perform a genetic screen to shortlist potential TF candidates that have novel roles in the development of the MTs.
- To choose the most interesting TFs to investigate and perform a detailed study of their possible novel functions in the MTs.
- To generate new methods and tools to identify the molecular mechanisms by which the selected TFs function.
- In parallel, to incorporate new ways of MT observation in the whole-fly.

The data and findings of these investigations will form the rest of this thesis and will provide novel insights into the development and homeostasis of the MTs of *Drosophila*. Thus, these results may also provide valuable information for the subsequent study of GATA factors in both invertebrate and vertebrate kidney.

Chapter 2 - Materials and Methods

2.1 *Drosophila melanogaster*

2.1.1 Rearing conditions

Drosophila were maintained in food vials with standard *Drosophila* medium (see Appendix 1) in ventilated capped tubes. The stocks were maintained at 22 °C, 55% atmospheric humidity and a 12:12 h light: dark cycle and adult flies were transferred to new tubes every two weeks. To obtain a progeny of a specific genotype, a minimum of 20 flies were transferred to a tube in a proportion of 3 virgin females per 1 male, and they were kept at higher temperatures (29 °C for the crosses using RNAi constructs, and 26 °C for other crosses). Adult progeny was collected just after emerging (0 to 1 day-old). All the experiments were performed with 5 to 7 days old flies unless specified otherwise.

2.1.2 Fly stocks

Genotype, purpose and origin of the fly lines used are summarised in Table 2.1-1.

Name/ID	Genotype	Purpose	Source/References
Canton-S	w ⁺ ;+/+;+/+	Wildtype strain used to outcross all the lines. Used for controls	Bloomington <i>Drosophila</i> Stock Center
<i>Bun</i> RNAi	y ^v ;+/+;P{TRiP.JF0295 4}attP2	UAS line encoding for an RNAi for <i>bun</i> mRNA, used for the screening	Bloomington <i>Drosophila</i> Stock Center, BL28322
<i>Cnc</i> RNAi	y ^v ;+/+;P{TRiP.JF0200 6}attP2	UAS line encoding for an RNAi for <i>cnc</i> mRNA, used for the screening	Bloomington <i>Drosophila</i> Stock Center BL25984
<i>Sv</i> RNAi	y ^v ;+/+;P{TRiP.JF0258 2}attP2	UAS line encoding for an RNAi for <i>sv</i> mRNA, used for the screening	Bloomington <i>Drosophila</i> Stock Center BL27269
<i>Ptx1</i> RNAi	y ^v ;+/+;P{TRiP.HMJ22 073}attP40	UAS line encoding for an RNAi for <i>ptx1</i> mRNA, used for the screening	Bloomington <i>Drosophila</i> Stock Center BL58124
<i>Corto</i> RNAi	y ^{sc^v} ;+/+;P{TRiP.GL003 82}attP2	UAS line encoding for an RNAi for <i>corto</i> mRNA, used for the screening	Bloomington <i>Drosophila</i> Stock Center BL35456
<i>HNF4</i> RNAi	y ^v ;+/+;P{TRiP.JF0253 9}attP2	UAS line encoding for an RNAi for <i>HNF4</i> mRNA, used for the screening	Bloomington <i>Drosophila</i> Stock Center BL29375

<i>STAT92e</i> RNAi	y ^v ;+/+;P{TRiP.JF0126 5}attP2	UAS line encoding for an RNAi for <i>STAT92e</i> mRNA, used for the screening	Bloomington <i>Drosophila</i> Stock Center BL31317
<i>Ada2B</i> RNAi	y ^{sc} *v ⁻ ;+/+;P{TRiP.GL002 43}attP2	UAS line encoding for an RNAi for <i>Ada2B</i> mRNA	Bloomington <i>Drosophila</i> Stock Center BL35334
<i>Smox</i> RNAi	y ^v ;+/+;P{TRiP.GL014 76}attP2	UAS line encoding for an RNAi for <i>smox</i> mRNA, used for the screening	Bloomington <i>Drosophila</i> Stock Center BL42318
<i>Hth</i> RNAi	y ^{sc} *v ⁻ ;+/+;P{TRiP.HMS01 112}attP2	UAS line encoding for an RNAi for <i>hth</i> mRNA, used for the screening	Bloomington <i>Drosophila</i> Stock Center BL34637
TM3/TM6	+/+;+/+TM3Sb/TM6 ,Tb,Hu	chromosome III balancer line	
<i>CapaR</i> -Gal4	w ⁻ ;+/+;w*, <i>CapaR</i> -Gal4	Gal4 under <i>CapaR</i> expression mainly expressed in the PCs.	In-house line (Terhzaz et al., 2012)
<i>CapaR</i> -Gal4:UAS GFP	w ⁻ ;/CapaRGal4;UAS.C D8.GFP	Gal4 under <i>CapaR</i> expression, with cytoplasmatic GFP expression, used to track <i>CapaR</i> expression and for OPT experiments.	In-house line (Terhzaz et al., 2012)
<i>CtB</i> -Gal4	P{ct-GAL4.CtB}	Gal4 under <i>Cut</i> expression, mainly expressed in the PCs, from the very early embryonic stages	Dr Barry Denholm (Saxena et al., 2014)
<i>Clc-a</i> -Gal4	w ⁻ ;+/+;w*, <i>Clc-a</i> -Gal4	Gal4 under <i>Clc-a</i> expression mainly expressed in the SCs, weaker expression in bar cells	Bloomington <i>Drosophila</i> Stock Center
<i>CapaR</i> ^{ts}	<i>CapaR</i> -Gal4;TubulinGal80. TS	<i>CapaR</i> Gal4 line thermosensitive, used for controlled time expression experiments	In-house line
<i>c42</i> -Gal4	w [*] ; P{w[+mW.hs]GFPBl o=GawB}c42	Gal4 expressed in the PCs	<i>Drosophila</i> Bloomington Stock Center, BL30835
<i>UrO</i> >GFP	w ⁻ ;UAS.CD8.GFP; <i>UrO</i> Gal4	Gal4 under <i>Urate Oxidase</i> expression expressed in a subset of PCs of the main segment.	In-house line (Terhzaz et al., 2010)
<i>GATAe</i> RNAi	w ¹¹¹⁸ ; P{GD4152}v10420;+ /+	UAS line encoding for an RNAi for <i>GATAe</i> mRNA, all isoforms, used for the main experiments with <i>GATAe</i>	Vienna <i>Drosophila</i> Resource Center v10420(Takashima et al., 2013)
<i>GATAe</i> RNAi (2)	y ^{sc} *v ⁻ ;+/+; P{TRiP.HMS01087}a ttP2	UAS RNAi line for <i>GATAe</i> mRNA, isoform A, used to confirm the phenotypes found with <i>GATAe</i> knockdown.	Bloomington <i>Drosophila</i> Stock Center (Okumura et al., 2016) BL33478
<i>GATAe</i> ⁻	w [*] ; P{ry ⁺ t7.2=neoFRT}82 B	Null mutant line for <i>GATAe</i>	Prof. Takashi Adachi-Yamada (Okumura et al., 2016).

	$TI\{w^{+mC}=TI\}GATAe^1$ / $TM3, Sb^1 Ser^1$		
<i>Esg>GFP^{ts}</i>	$w^*; esgGal4.UAS.GFP/Cyo; tubGal80ts/TM6B$	Gal4 under <i>escargot</i> expression, in the RNSCs, thermosensitive, used for controlled time expression experiments	Dr Julia Cordero (Tian et al., 2017)
UAS-GATAe-HA	y(Consortium) w[67c23]; P{y[+t7.7] pUAST-CFLAGHA-BD-PHI UAS.GATAe-HA, w ⁺ =CaryP}attP2	UAS line expressing GATAe isoform A with an HA tag in the C-terminal, used for overexpression and ChIP-seq experiments	Guillermo Martinez Corrales. generated in this study.
GATAe.TRiP/TM3	$y^1 sc^+ v^1; P\{TRiP.HMS01253\}attP2/TM3, Sb^1$	UAS line encoding for an RNAi for GATAe mRNA, all isoforms, used to confirm the phenotypes found with GATAe knockdown	Bloomington <i>Drosophila</i> Stock Center (BL34907)
<i>Actin-Gal4/Cyo GFP</i>	y1w-; P{act5cGAL4}25FO1/Cyo-GFP, y+; +; +	Gal4 under <i>Actin</i> expression, expressed ubiquitously, balanced with GFP. Also used for balancing other lines with Cyo/GFP	No longer in Bloomington but kept in-house.
TM3/UAS.GFP	$w^*; Sb^1 / TM3, P\{w^{+mC}=ActGFP\}JM R2, Ser^1$	GFP balancer in the chromosome III.	Bloomington <i>Drosophila</i> Stock Center BL4354
Sco	$w^*; Sco/Cyo; TM6/M KRS$	Double balancer in the Chromosomes II and III	Bloomington <i>Drosophila</i> Stock Center.
Ras ^{N17}	$P\{UAS-Ras85D.N17\}TL1, w^{1118}$	UAS line expressing a dominant negative (DN) form of <i>Ras85D</i>	Bloomington <i>Drosophila</i> Stock Center BL4846, (Lee et al., 1996)
Ras ^{V12}	$w^{1118}; +/+; P\{UAS-Ras85D.V12\}TL1$	UAS line expressing a constitutively activated form of <i>Ras85D</i>	Bloomington <i>Drosophila</i> Stock Center BL4847, (Wu et al., 2010).
GATAe-Gal4 [VT04235 7]		Gal4 controlled by GATAe regulatory sequences.	Vienna <i>Drosophila</i> Resource Center v209818 (Zhai et al., 2018).
GATAe-Gal4 [VT04235 8]		Gal4 controlled by GATAe regulatory sequences.	Vienna <i>Drosophila</i> Resource Center v205732 (Zhai et al., 2018).

Table 2.1-1. List of *Drosophila* stocks used in this study including the stock name, genotypes, the purpose and the origin of the stocks.

2.1.3 Embryo collection

Approximately 50 females and 15 males were pooled in an inverted plastic beaker, which has both ends open. One of the ends is covered with a fine-mesh to allow air to flow. Flies were allowed to lay eggs on a plastic 55mm Petri dish containing grape juice agar medium (see Appendix 2) with a dab of yeast paste (dried yeast dissolved in double-distilled water, ddH₂O). Flies were allowed to lay overnight at 29°C for Gal4/UAS experiments, or 22°C for other experiments on the plate unless specified otherwise, and the plate was removed the following morning for embryo collection. Embryos were detached from the egg-laying plate using a paintbrush and a stream of distilled water for subsequent fixation and collected in a fine-mesh basket.

2.1.4 Dissection of *Drosophila* MTs

7-day old adult female flies were anaesthetised on ice and used for tissue dissection (except for flies kept at 29°C, which were 5-day old flies). MTs were dissected in ice-cold *Drosophila* Schneider's Medium (Invitrogen), and transferred to Poly-D-lysine treated microscope slides (Thermo Fisher Scientific), containing 100µl of PBS (Phosphate buffered saline, pH 7.4: 137 mM NaCl, 2.7 mM, KCl, 10 mM Na₂HPO₄, 1.8 mM KH₂PO₄) for subsequent immunocytochemistry (ICC) experiments (section 2.6). Otherwise, dissected tubules were transferred every 10min to an Eppendorf tube containing the appropriate solution for Western-Blot, RNA extraction, etc. Third instar larvae (L3) and pupae, were placed in aluminium foil on ice before dissection, to avoid movement of the L3 specimens.

2.2 *Escherichia coli*

2.2.1 Strains, Vectors and Transformation conditions

The plasmids (Table 2.2-1) were transformed into DH5α® subcloning efficiency competent cells (Invitrogen) by adding 100ng of the plasmid to 100µl of cells and incubated on ice for 30 min. Cells were then heat shocked at 42°C for 45 secs with no shaking. Then they were left on ice for 2 minutes before adding 250µl of room temperature (RT) S.O.C medium® (Invitrogen, Appendix 3), and incubated at 37°C for 1h, shaking. Then, 100µl of the transformation were spread onto L-

agar (see Appendix 3) plates containing 75µg/ml carbenicillin, and kept overnight at 37°C. The positive colonies, carbenicillin resistant, were selected.

Vector	Purpose and origin
pUAST-CFLAGHA-BD-PHI GATAe isoform A driven by a UAS promoter and with a Flag-HA tag C-terminal fusion.	Used for generation of an HA-tagged form of GATAe, E. coli expression experiments, <i>Drosophila</i> S2 cells transformation and <i>Drosophila</i> embryo microinjection (Clone ID UFO01424, <i>Drosophila</i> Genomics Resource Center).
pUC vector	Negative control for E. coli expression experiments (NEB).

Table 2.2-1. Vectors used in this study.

A gain-of-function line for *GATAe* has been generated in this project, from a plasmid obtained as a bacterial slant from the *Drosophila* Genomics Resource Centre (DGRC, clone UFO01424). The resulting protein is tagged with Human influenza hemagglutinin (HA) in the C-terminal (UAS-*GATAe*-HA, Figure 2.2.1). Plasmid DNA was isolated as described in 2.2.2. The UAS-*GATAe*-HA plasmid contained both ampicillin/carbenicillin and chloramphenicol resistance genes. The antibiotic carbenicillin was used in this study for plasmid selection. The stock solution was made by preparing a stock solution of 7.5mg/ml of carbenicillin in ethanol absolute, stored at -20°C. Then it was diluted in 1/100 in the agar broth before preparing the plates. The growing colonies, carbenicillin resistant were selected and resuspended in 10µl of ddH₂O.

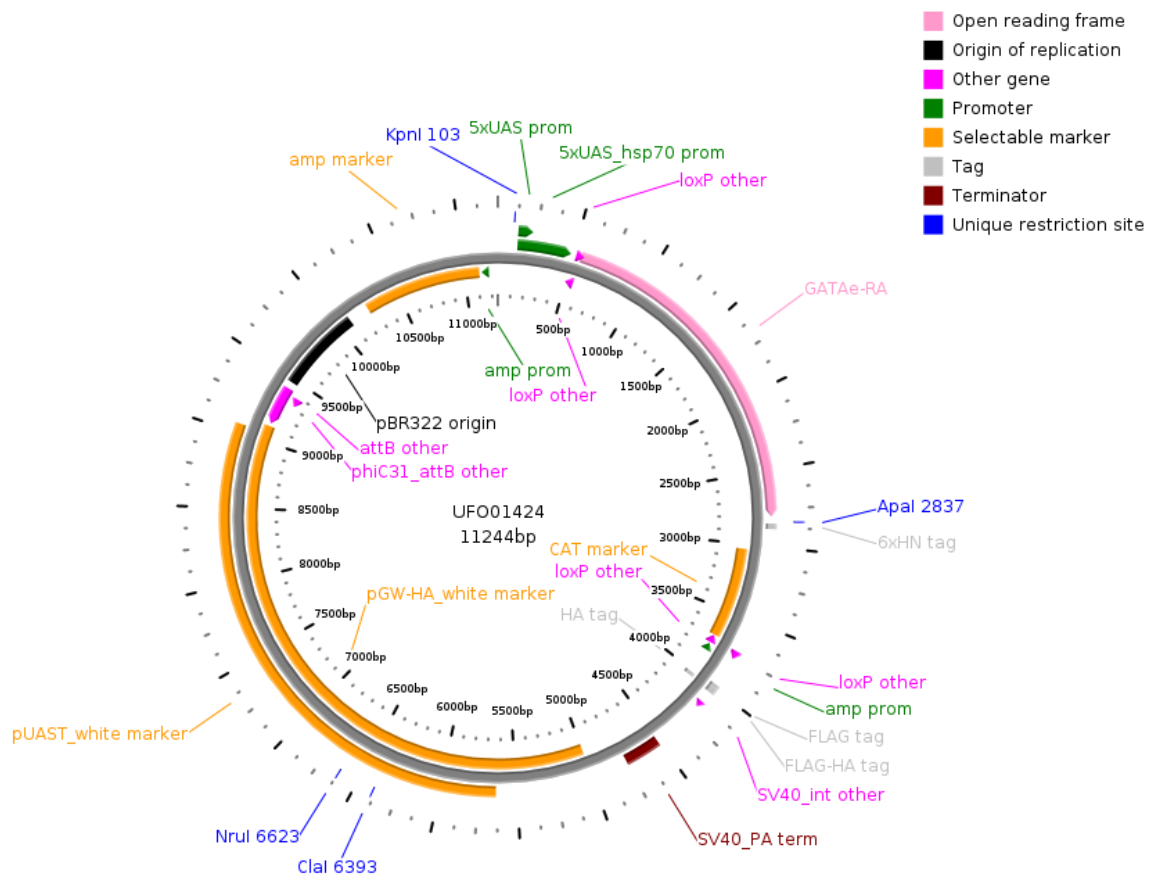


Figure 2.2.1. Schematic map of the pUAST vector, used for *GATAe* overexpression. It includes a FLAG-HA-tag sequence in the C terminal of the gene sequence.

2.2.2 Isolation of plasmid DNA

The carbenicillin resistant colonies were resuspended in 10µl of ddH₂O, added to 5ml of LB and incubated overnight at 37°C shaking. Plasmid DNA was extracted using the QIAprep Spin MiniPrep Kit (Qiagen), following manufacturer's instructions, and eluted in 30µl of elution buffer. DNA was quantified using a NanovuePlus® Spectrophotometer.

2.2.3 Verification of plasmid insertion

PCR experiments were performed to check the *GATAe* positive colonies using primers against *GATAe* ORF sequence (Table 7.3-2). The purified plasmid DNA was cut using the specific restriction enzymes KpnI, NheI and BamHI and run in a 1% agarose gel to verify the size and number of fragments. Finally, the isolated DNA was sent for sequencing to confirm *GATAe* sequence integrity, using primers against *GATAe* sequence. Sequencing was performed by GATC Biotech, UK. and subsequently analysed using Mac Vector Software. Bacteria containing the proper *GATAe* plasmid were stored in glycerol stocks (2% w/v peptone, 40% v/v glycerol

solution in H₂O) at -80°C. To recover the frozen stock, a small amount was scraped with a pipette tip, and inserted into a tube containing 5ml of LB with the appropriate antibiotic to grow at 37°C overnight. Confirmation of gene expression and protein localisation was performed by co-transfection of UAS-*GATAe*-HA plasmid and metallothionein-induced promoter pMT-Gal4 (Santos et al., 2007). UAS-*GATAe*-HA plasmid was sent for microinjection in *Drosophila* embryos using the Gene® microinjection services. Five different lines were obtained and were studied for morphological and functional phenotypes.

2.3 *Drosophila* S2 Cells

2.3.1 Maintenance

Drosophila Schneider 2 (S2) cells (Invitrogen) were kept as described in the manufacturer's instructions. All procedures were performed in sterile conditions, in a BioMAT² safety cabinet. Cells were maintained in Complete Serum Medium (CSM), composed of 10% of Heat Inactivated Fetal Bovine Serum (FBS, Sigma) in *Drosophila* Schneider's Medium (GIBCO™). Cells were kept in Corning® 75cm² Cell Culture Flasks, at 28°C in a Cell Culture Heated Incubator (PHCbi). Cells were passaged every three days (when they reached a concentration of 10⁷ cells/ml) by diluting 2ml of cells in 13ml of fresh CSM in a new flask.

2.3.2 Transient transfection

S2 cells were transferred to 6-well plates for transient transfection when they reached a concentration of 1x10⁶ cells/ml. 3ml of cells were co-transfected with both pMT-Gal4 and UAS-*GATAe*-HA plasmids using an Effectene® Transfection Reagent following manufacturer's instructions. Cells were then kept at 28°C overnight, and copper sulphate (CuSO₄) was added at a final concentration of 500µM. Cells were then incubated for two days to allow protein expression. Cells were resuspended in the medium and transferred to a 15ml capped tube and centrifuged at 1500g for 2min at RT to be harvested. The cells were then washed twice with PBS, and kept at -80°C.

2.4 Molecular biology protocols

2.4.1 RNA extraction

2.4.1.1 Adult tubules

For RNA extraction of tubules, a total of 50 adult female flies were dissected in Schneider's *Drosophila* Medium (GIBCO™) on ice and immediately transferred to a 1.5mL Eppendorf tube containing 250µl of RLT buffer + 1% β-mercaptoethanol. Samples were subjected to short bursts of sonication using a Microson™ Ultrasonic Cell Disruptor. Following this, RNA was extracted using a Qiagen® RNeasy kit, following manufacturer's instructions. RNA was eluted with 30µl of nuclease-free ddH₂O and quantified using a NanovuePlus® Spectrophotometer.

2.4.1.2 Whole fly

For RNA extraction of whole flies, a total of 5 adult female flies were anaesthetised on ice, immediately transferred to a 1.5mL Eppendorf tube containing 350µl of QIAzol (Qiagen) and homogenised with a clean plastic rod. 70µl of chloroform was added, and the solution was vortexed for 30sec and let stand for 3min. Samples were centrifuged at 13000 revolutions per minute (rpm) at 4°C for 5min in an AccuSpin™ Micro R Ultra Centrifuge. After removing the organic phase, the samples were spun again at 13000 rpm, 4°C for 5min and the aqueous part was placed in a new clean tube. Then, 1.5 volumes of 100% ethanol were added to the samples and RNA was extracted using a Qiagen® RNeasy kit, following the manufacturer's instructions. RNA was eluted with 30µl of nuclease-free ddH₂O and quantified using a NanovuePlus® Spectrophotometer.

2.4.2 cDNA synthesis

Complementary DNA (cDNA) was synthesised from 500ng of RNA using SuperScript II reverse transcriptase (SSII, SuperScript® II, 200 units/µl, Invitrogen), following manufacturer's instructions. 1µl of OligoDT₁₂₋₁₈(500µg/ml), 1 µl of dNTP mix (10mM), 4 µl of 5X First Strand Buffer, 2 µl of DTT (0.1M) and 1 µl of RNase Out inhibitor (40 units/µl, Invitrogen) were combined with 500ng of total RNA and the mix made up with nuclease-free ddH₂O to obtain a volume of 19 µl. The mix was briefly centrifuged and incubated for 5min at 65°C using a

PCR block. Samples were chilled on ice for 1min, and 1 µl of SSII was added, following quick centrifugation. Samples were incubated at 42 °C for 1h and 70 °C for 15min. cDNA was stored at -20 °C. To control for any genomic amplification, a control tube for each sample was prepared without SSII. cDNA was quantified using a NanovuePlus® Spectrophotometer.

2.4.3 Polymerase chain reaction (PCR)

2.4.3.1 Standard PCR

PCR was performed using DreamTaq® Green PCR MasterMix (Thermo Fisher). The PCR master mix contained 10µl of PCR MasterMix (DreamTaq DNA Polymerase, DreamTaq buffer, 4mM MgCl₂ and 0.4mM of each dNTP), 1µl of each forward and reverse primer, 6µl of nuclease-free dd H₂O and 2µl of cDNA, and the thermal cycling conditions are summarised in Table 2.4-1. PCR products were analysed using agarose gel electrophoresis described in (2.4.5). In order to test the efficiency of the primers (of both PCR and qPCR), gradient PCR with DreamTaq enzyme was performed, and the optimal annealing temperature was established for each primer pair.

Step	Number of Cycles	Temperature (°C)	Time
<i>Initial denaturation</i>	1	95	2min
<i>Denaturation</i>	35	95	30sec
<i>Annealing</i>		55-60 (T _m -5)	30sec
<i>Extension</i>		72	1min/kb
<i>Final extension</i>	1	72	5min

Table 2.4-1. Normal PCR thermal cycling conditions.

2.4.3.2 qPCR

Quantitative PCR (qPCR) was used to determine the expression of a gene of interest. In dye-based qPCR, fluorescent labelling allows the quantification of the amplified DNA molecules by employing the use of a dsDNA binding dye. During each cycle, the fluorescence is measured. The fluorescence signal increases proportionally to the amount of replicated DNA, and hence the DNA is

quantified in “real time”. The disadvantages to dye-based qPCR are that only one target can be examined at a time and that the dye will bind to any ds-DNA present in the sample. Each qPCR reaction was composed by 10µl of Agilent 2X Brilliant III SYBR® Green QRT-PCR Master Mix (Agilent Technologies), 1µl of each forward and reverse primers, 6 µl of nuclease-free ddH₂O, and 2µl of cDNA. The solution was gently mixed and briefly centrifuged before running the protocol, summarised in Table 2.4-2. The melting curve was analysed, and if any of the curves produced more than one peak, primers were discarded. Data were analysed using StepOne™ v2.1 Software and plotted in a graph using GraphPad Prism® 5 Software. Statistical significance was assessed using the 1-way ANOVA test. Relative quantification was calculated comparing the ratio of DNA concentration of the gene of interest with the ratio of DNA concentration of the housekeeping gene *rpl32*, using the C_T method (Schmittgen and Livak, 2008). All primer sequences used in this study can be found in Appendix 4.

Step	Number of Cycles	Temperature	Time
<i>Initial denaturation</i>	1	95 °C	10min
<i>Denaturation</i>	40	95 °C	15sec
<i>Annealing,</i>		55-60 (T _m -5)	30sec
<i>Extension and fluorescence reading</i>		80 °C	10sec
<i>Melting curve stage</i>	1	95 °C	15sec
	1	60 °C	1min
	1	95 °C	15sec

Table 2.4-2. qPCR thermal cycling conditions.

2.4.4 Restriction enzyme digestion

Restriction enzyme digestions were performed to verify the size of the plasmid and digested fragments. Restriction enzyme sequences in the UAS-GATAe-HA plasmid were identified using the online software PeptideCutter (Gasteiger et al., 2005). All digests were performed using New England Biolabs (NEB) enzymes and recommended buffers for each enzyme or combination of enzymes, using

manufacturer's instructions. To confirm if the DNA fragments were of a correct size, digestions were run on a 1% agarose gel.

2.4.5 Agarose gel electrophoresis of DNA

Different DNA sizes were separated in a 1% agarose gel (containing 0.1µl/ml of ethidium bromide), dissolved in 0.5X TBE buffer (see Appendix 5). Before loading into the gel, 5X GelPilot® Loading Dye was added in a ratio of 1/5 to each DNA sample, except the PCR products obtained using DreamTaq® Green PCR MasterMix. DNA was visualised under ultraviolet light, and the size of DNA fragments was compared to a 100bp or 1kb DNA ladder (Invitrogen).

2.4.6 Western blotting

2.4.6.1 Preparation of the samples

50 adult female flies (100 pairs of tubules) were dissected in ice-cold Schneider's *Drosophila* Medium (GIBCO™) for each sample and immediately transferred to a 1.5mL Eppendorf tube containing 50µl of ice-cold RIPA buffer (Thermo Scientific) complemented with 1X proteinase inhibitor cocktail (Sigma) and kept on ice. The lysate was then sonicated on ice and stored at -80°C until further use.

2.4.6.2 Bradford protein assay

Bradford assays were performed on a 96 well plate (Corning Inc.). Standards from 0-5 µg (0 µg, 0.5 µg, 1 µg, 2 µg, 3 µg, 4 µg, 5 µg) were generated in triplicate using Bovine Serum Albumin (BSA, New England Biolabs), in a volume of 50µl ddH₂O. 1µl of each sample was diluted into 49µl of ddH₂O in triplicate. Each well contained either the protein samples or the standard samples, 200µl of a well-mixed 1/5 dilution of Bio-Rad protein assay dye reagent concentrate (Bio-Rad) in ddH₂O. Absorbance at 590 nm was measured using a Mithras LB 940 plate reader (Berthold technologies); Microsoft Excel software was used to generate a standard curve and calculate the concentration for each sample.

2.4.6.3 Running Gel electrophoresis

A total of 20µg of lysate was diluted in ice-cold RIPA buffer, up to a volume of 20µl, and 5µl of NuPAGE® LDS 4X Sample Buffer (Invitrogen). 2.5% (v/v) of β-Mercaptoethanol was added to each sample and then heated at 70°C for 10 min to reduce the sample. Electrophoresis was performed in a NuPAGE™ 4-12% Bis-Tris Gel (Invitrogen), using an XCell SureLock™ Electrophoresis System (Invitrogen), and gels were run at 200V for 1h on ice in 1X NuPAGE™ MES SDS Running Buffer (Life Technologies).

2.4.6.4 Transfer Blotting

Immobilon®-P Transfer Membrane was activated by incubation 2min in 100% methanol and washed with ddH₂O. Protein transfer was performed using an XCell SureLock™ Electrophoresis System, in Transfer Buffer (5% methanol, 5% NuPAGE™ Transfer Buffer (Life Technologies), in ddH₂O) at 100V for 1h on ice.

2.4.6.5 Ponceau Staining

Ponceau staining method was used to visualise the protein present in the membrane. The membrane was washed with TBST (TBS 1X + 0.2%Tween, see Appendix 6). It was incubated with Ponceau solution (1% Ponceau S, 3% TCA in ddH₂O) for 5min, and washed twice with 10% glacial acetic acid for 5 min each wash.

2.4.6.6 Antibody Incubation

The membrane was washed three times with TBST for 10 min each and was incubated with TBST + 5% Skimmed Milk blocking solution for 30 min. Primary antibody or combination of antibodies was added at the required concentrations (Table 2.6-1) at 4°C overnight. The membrane was then washed three times with TBST and once with TBST + 5% Skimmed Milk blocking solution, following incubation of the HRP-conjugated secondary antibody (Table 2.6-2) for 1h at RT and in the dark. Finally, the membrane was extensively washed with TBS for 1h and prepared for ECL detection.

2.4.6.7 Membrane ECL Detection

Chemiluminescence detection was performed using the ECL® Western Blotting analysis system (Amersham Pharmacia) following the manufacturer's instructions. Equal volumes of solution 1 and solution 2 (see Appendix 7 for the recipes) were mixed and added to the membrane in a sheet of plastic wrap and incubated for 2min. The membrane was dried with paper and wrapped in plastic cling film, added to a cassette, and exposed to a Medical X-ray Blue/MXBE ECR Film (Carestream), and developed in an X-OMAT film processor.

2.5 Generation of the antibody against GATAe protein

2.5.1 Antigenic peptide design

The optimal immunogenic sequence from the GATAe protein sequence was identified using the online software AbDesigner (Pisitkun et al., 2011) as **SPYSQQSTPQSQSPH**. A cysteine amino acid was added to the C-terminal of the sequence to increase the sequence immunogenicity. Then, the peptide was sent to Genosphere Biotechnologies to be injected in rabbits. The resulting lyophilised powder obtained from the rabbit serum was resuspended in ddH₂O and peptide-purified using a HiTrap NHS-Activated HP affinity column (GE Healthcare Life Sciences), following the manufacturer's instructions. The specificity of the antibody was tested using immunocytochemistry and Western Blot techniques.

2.6 Immunofluorescence

2.6.1 Immunofluorescence of larval, pupal and adult tubules

All samples were dissected in ice-cold PBS and transferred to Poly-lysine treated microscope slides and fixed in 4% paraformaldehyde (PFA) for 20min at RT, except for the ones stained with DCAD antibody, which were fixed during 10min. Samples were washed 3 times with PBTA (1xPBS, 0.5% Triton-X100, 0.2% Sodium Azide), 20min each, and incubated for 2h with PBTA-NGS blocking solution (PBTA, 10% Normal Goat Serum, Sigma-Aldrich) at RT. Samples were incubated at 4°C overnight in PBTA-NGS blocking solution with the appropriate primary combination of antibodies (Table 2.6-1). Tissues were then washed 5 times with

PBTA, 20min each, and blocked with PBTA-NGS for 2h at RT. Samples were then incubated with the appropriate secondary antibodies in PBTA-NGS for 2h at RT, in the dark (Table 2.6-2). Tissues were washed 3 times with PBTA and once more with 1xPBS, for 20min each, at RT and in the dark. When needed, samples were incubated with Phalloidin (3.6µl of stock Phalloidin solution in 500µl of PBS) for 30min in the dark, followed by 3 washes with PBS, 20min each at RT, in the dark. They were then incubated for 2minutes with DAPI and mounted with Vectashield® (VectorLabs) antifade mounting medium for fluorescence.

2.6.2 Immunofluorescence of Embryos

Embryos were collected from 10 to 16h after egg laying, dechorionated with 100% bleach for exactly 2 min, washed with ddH₂O and then fixed in 4% PFA/heptane for 20 min at RT, except for the ones stained with DCAD antibody, which were fixed during 10 min in rotation and at RT. Vitelline membranes were removed by washing 3 times with 100% methanol (Thermo Fisher). Embryos were washed 3 times with PBTA, for 20min each, and were incubated with PBT-NGS blocking solution. Embryos were incubated at 4°C overnight in PBT-NGS with the appropriate primary antibodies (Table 2.6-1). Embryos were washed 5 times with PBTA, 20 min each, and blocked with PBT-NGS for 2h. They were incubated with the appropriate secondary antibodies (Table 2.6-2) in PBT-NGS for 2h at RT, in the dark. They were washed 3 times with PBTA and once more with 1xPBS, for 20 min each, at RT and in the dark. The embryos were mounted in Vectashield® in microscope slides.

2.6.3 Immunofluorescence of S2 cells

Co-transfected S2 cells were collected from the flasks at a concentration of 5×10^7 cells/ml and resuspended in a 15ml tube. 100µl of cells were taken and then allowed to settle in poly-lysine treated microscope slides for 10min at RT. Once the cells attached to the slides, S2 cells were fixed in 4% PFA for 10 min at RT. The number and length of washes of the samples were the same as the performed with embryos (see 2.6.2). After the last wash with 1xPBS, S2 cells were incubated for 2 minutes with DAPI and mounted with Vectashield® (VectorLabs).

2.6.4 List of primary and secondary antibodies used

Antibody	Host species and Dilution	Purpose	Source and reference
Anti-ct	Mouse, 1/100	Monoclonal, against Cut protein, present in the nuclei of PCs	Developmental Studies Hybridoma Bank (2B10)
Anti-CLC-a	Rabbit, 1/100	Polyclonal, against CLC-a protein, present in the basal membrane of SCs	(Cabrero et al., 2014)
Anti-Dlg	Mouse, 1/500	Monoclonal, against Discs Large protein, present in the basolateral membrane	Developmental Studies Hybridoma Bank (4F3)
Anti-DL	Mouse, 1/50	Monoclonal, against Delta protein, Stem Cell marker, a component of Notch Signalling	Developmental Studies Hybridoma Bank (C594.9B)
Anti-arm	Mouse, 1/100	Monoclonal, against Armadillo protein, in the membranes of RNSCs	Developmental Studies Hybridoma Bank (N27A1)
Anti-GFP	Rabbit, 1/1000	Polyclonal, against Green Fluorescent Protein	Life Technologies #A-11122
Anti-PH3	Rabbit, 1/1000	Polyclonal, against Phospho-histone 3, dividing cells marker	Thermo Fisher Scientific
Anti-DCAD	Rat, 1/200	Monoclonal, against the extracellular domain of De-cadherin, adherens junctions and marker of RNSCs	Developmental Studies Hybridoma Bank (DCAD2)
Anti-HA	Rat, 1/100 (1/2000 in Western Blot)	Monoclonal, against peptide sequence (YPYDVPDYA), derived from the influenza hemagglutinin protein	Thermo Fisher (3F10)
Anti-Hnt	Mouse, 1/20	Monoclonal, against Hindsight protein, marker for RNSCs	Developmental Studies Hybridoma Bank (1G9)

Table 2.6-1. List of primary antibodies used in this project.

Antibody	Host species and Concentration	Purpose	Source and reference
Anti-mouse-IgG-488	Goat, 1/1000	Polyclonal, against mouse IgG, conjugated to AlexaFluor 488	Life Technologies
Anti-mouse-IgG-546	Goat, 1/1000	Polyclonal, against mouse IgG, conjugated to AlexaFluor 546	Life Technologies
Anti-mouse-IgG-633	Goat, 1/1000	Polyclonal, against mouse IgG, conjugated to AlexaFluor 633	Life Technologies
Anti-rabbit-IgG-488	Goat, 1/1000	Polyclonal, against rabbit IgG, conjugated to AlexaFluor 488	Life Technologies
Anti-rabbit-IgG-546	Goat, 1/1000	Polyclonal, against rabbit IgG, conjugated to AlexaFluor 546	Life Technologies
Anti-rat-IgG-488	Goat, 1/1000	Polyclonal, against rat IgG, conjugated to AlexaFluor 488	Life Technologies
Anti-rat-IgG-H&L (HRP)	Goat, 1/2000 (Western Blot)	Polyclonal, against rat IgG, conjugated to Horse Radish Peroxidase (HRP), used for Western Blot	Abcam (ab97057)

Table 2.6-2. List of secondary antibodies used in this project.

Marker	Concentration	Purpose	Source
DAPI	1/10000	4',6-Diamidino-2-Phenylindole, Dihydrochloride, stains nuclei	Sigma Aldrich
Phalloidin-TRITC	7.2µl/ml of PBS	TRITC- (tetramethylrhodamine B isothiocyanate) conjugated Phalloidin (F-Actin specific fungal toxin).	Sigma Aldrich

Table 2.6-3. List of other staining markers used.

2.6.5 Whole fly imaging sample preparation

2.6.5.1 Adult fly preparation

CapaR Gal4>UAS-GFP (*CapaR*>GFP) were used for experiments in which the MTs were observed in whole-fly preparations. Adult flies were kept at 29°C for 10 days to obtain a maximum GFP signal in the MTs. Flies' legs, wings and heads were removed, and bodies were fixed with 8% PFA for 24h. The abdominal cuticle was carefully removed manually without disturbing the internal organs' organisation, and sample was incubated with CUBIC clearing solution (see Appendix 8) for 24-48h (until the specimen became transparent) in the dark, in rotation and at RT. Once the sample became transparent, it was glued in a pipette tip, allowing complete 3D rotation, for posterior scanning using OPT (Figure 2.7.1).

2.6.6 Larvae preparation

CapaR>GFP wandering L3 specimens were selected by their GFP expression and collected from the tube. They were incubated with 8% PFA for 24h and carefully pierced several times to allow the fixative agent to reach all the internal organs. Samples were incubated with CUBIC clearing solution for 24-48h (until the sample is transparent) and mounted in a 55mm petri dish containing a drop of 1% warm agarose. Once the agarose solidified, samples were scanned using a two-photon microscope (Figure 2.7.2).

2.6.6.1 Embryo preparation

Embryos were collected, fixed and immunostained with Ct antibody (Table 2.6-1) as described in 2.6.2 to stain their MTs. They were mounted in a 55mm petri dish containing a drop of 1% warm agarose. Once the agarose solidified, the embryos were scanned using a multiphoton microscope (Figure 2.7.2).

2.7 Imaging

2.7.1 Confocal Microscopy

Immunostained S2 cells and Malpighian tubules were examined using a Zeiss LSM 880 Axio Observer microscope (Zeiss, UK) using 10x, 20x and 63x (oil immersion)

and processed with Zeiss Blue software and Adobe Illustrator CS6. DAPI was excited at 405 nm wavelength and emitted at ~420 nm. AlexaFluor-488-conjugated secondary antibodies were excited at 488 nm and emitted at ~520 nm. TRITC-conjugated phalloidin and the Alexa Fluor 546-conjugated secondary antibodies were excited with 543 nm line and emitted at ~575 nm. For multiple fluorescence imaging, each of the fluorophores was sequentially scanned to excite each fluorophore individually to avoid cross-reaction between channels. Scanned images were imported to Zeiss Blue Software® for processing, and final figures were produced using Adobe Illustrator CS6® Software.

2.7.2 Optical Tomography Projection Microscopy

For Optical Tomography Projection Microscopy (OPT), a Bioptonics 3001 device was used. It has a resolution of 1024x1024 (3.1µm), and samples were analysed using bright field and GFP (425/40nm exciter and 475nm LP emitter) to visualise the GFP fluorescence localised to the Malpighian tubules. Data obtained from the scanner was further processed with the 3D analysis software and posteriorly processed/analysed with Imaris 7.4® Software, obtaining a clear 3D image of the sample, which was exported as a movie (.AVI).

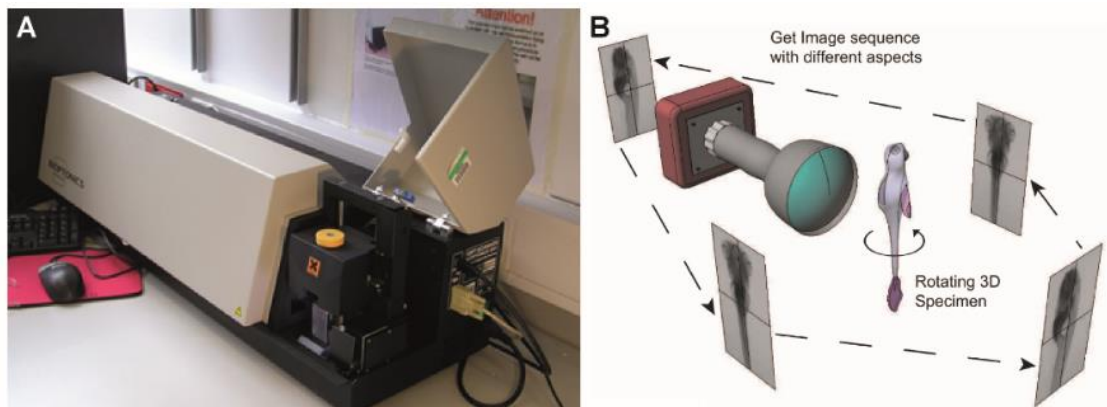


Figure 2.7.1. Schematic diagram of the OPT microscope.

(A) Picture of the Bioptonics 3001 device, used for OPT microscopy. Picture from <https://www.kcl.ac.uk/dentistry/research/divisions/craniofac/researchgroups/sharpe/label/optical-projection-tomography.aspx> (B) Schematic diagram of a sample 3D scanning of a specimen (in this study, a whole *Drosophila* adult fly). Pictures are taken on the rotating sample and then processed to obtain a 3D image of the whole sample. Taken from (Fei et al., 2012).

2.7.3 Two-photon Microscopy

For two-photon excitation microscopy, a two-photon Nikon A1R+MP microscope was used. This technology allows imaging of living and fixed samples up to 1mm

depth, to see the internal organs. Two-photon microscopes do not possess a pinhole, unlike the confocal microscopes. The concept of this technology is that two photons with lower energy (typically infrared wavelength, 700nm) than that required for a single photon, can excite a fluorophore similarly. Since the excitation only occurs when both photons coincide, photobleaching and background noise are strongly reduced compared to other systems. A schematic diagram of two-photon microscopy is shown in Figure 2.7.2. The samples were excited with 820nm-1300nm lasers and fixed 1040nm dual output femtosecond lasers, which allowed the observation of GFP⁺ structures. The objectives CFI75 Apochromat 25X W MP1300 (NA 1.10, WD 2.0 mm, water immersion), without a cover slide, and CFI PLAN FLUOR 20X MI (NA 0.75, WD 0.33-0.35mm, Water-Glycerin-Oil), with cover slide, were used. Samples were posteriorly analysed with NIS-Elements C-ER Software and Imaris 7.4® Software. 3D images were finally exported as pictures (.TIFF) or as a movie (.AVI).

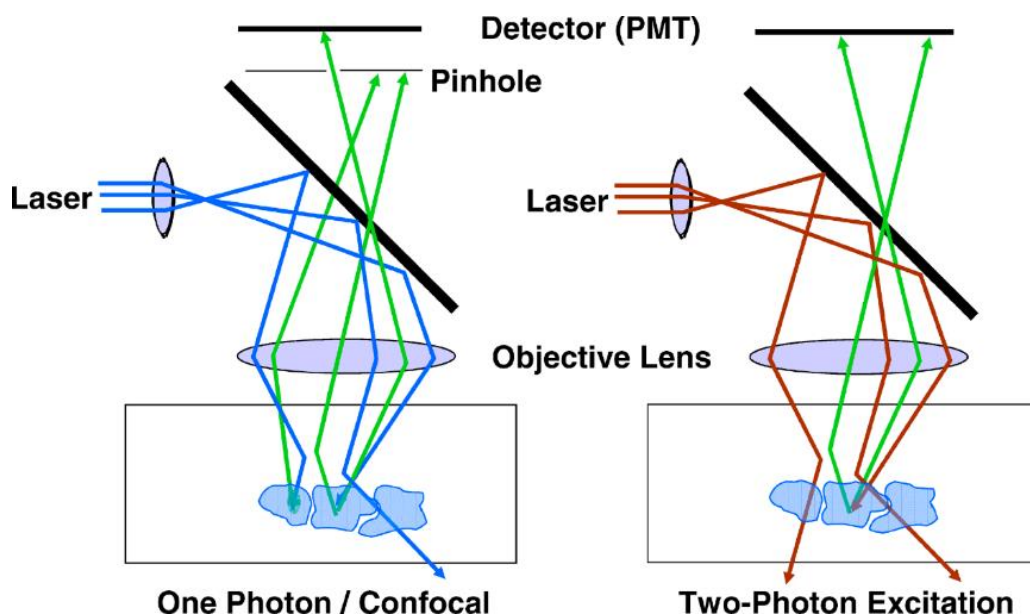


Figure 2.7.2. Schematic diagram comparing Two-Photon microscopy vs confocal microscopy.

Comparison between two-photon (right) and confocal microscopy (left). In comparison to confocal microscopy, Two-Photon microscopes do not have a pinhole, allowing all the excitation that comes from the sample to be collected. At the same time, because only when two photons excite one point at the same time, there is going to be excitation, the background and detection are improved, adapted from (Piston, 2005).

2.8 Chromatin immunoprecipitation sequencing

All Chromatin Immunoprecipitation (ChIP) experiments were performed using a SimpleChIP® Plus Enzymatic ChromatinIP Kit (*Cell Signaling Technologies*,

#9005). All reagents, unless specified otherwise, were obtained from this kit. The ChIP protocol provided in the kit was adapted as described below, and the original protocol from Cell Signalling Technologies can be found in <https://www.cellsignal.co.uk/products/chip-kits-reagents/plus-enzymatic-chromatin-ip-kit-magnetic-beads/9005>.

2.8.1 Sample collection

2.8.1.1 S2 cells collection

GATAe expressing S2 cells were maintained and co-transfected as described in 2.3. Initially, a total of 1×10^7 cells were used for each experiment. However, the number of cells was modified as showed in section 6.5.5. Cells were resuspended in 2ml of PBS with 1% Protease Inhibitor Cocktail (PIC) and fixed with 1% formaldehyde for 10 min at RT on a rotating wheel. This reaction was stopped by adding 200µl of 10X Glycine and mixing for 5 min at RT. Fixed cells were centrifuged at 500 x g for 5 min at 4°C, and pellets were resuspended in 2ml of PBS+PIC, three times. Cells were centrifuged at 2000 x g for 5 min at 4°C, and the supernatant was carefully removed. Supernatants were stored at -80°C until further use.

2.8.1.2 Whole fly collection

Total weight of 50 mg (different quantities of flies are commented in the discussion- include page no.) of adult *CapaR>GATAe*-HA flies for each ChIP were briefly anesthetised in CO₂ and transferred to a 2ml Eppendorf tube containing 1.8ml of PBS+PIC. They were homogenised using a Bead Ruptor 24 Elite, for 3 pulses of 30 sec, at a speed of 6.45 m/s. Between each pulse, samples were kept for 1min on ice to avoid overheating. Homogenates were fixed with 1.5% formaldehyde for 20 min at RT, with rotation. After stopping the reaction with 1X glycine and performing the same number of washes and centrifugations as with S2 cells (see 2.8.1.1), samples were then sonicated (5 pulses of 10 sec each) on ice using a Microson™ Ultrasonic Cell Disruptor to get a single-cell suspension. Tissue was now ready for nuclei preparation or stored at -80°C until further use.

2.8.1.3 Malpighian tubule collection

A total of 1000 5-to-7day-old adult *CapaR>GATAe*-HA flies (2000 pairs of tubules) were dissected in batches of around ~200 flies, collected in 1ml of PBS+PIC and were fixated independently. Fixation, glycine treatment, washes, centrifugations and sonication steps were the same as in the whole fly preparations (2.8.1.2). Tissue was now ready for nuclei preparation or stored at -80°C until further use.

2.8.2 Nuclei preparation

S2 Cells/whole fly tissue/tubules samples were resuspended in 1ml ice-cold 1X buffer A + PIC + 0.5mM DTT, and incubated on ice for 10 min, with tube inversion every 3 min. Homogenates were centrifuged at 2000 x g for 5 min at 4°C, and the pellets were resuspended with 1ml ice-cold buffer B + 0.5mM DTT. After centrifugation at 2000 x g for 5 min at 4°C, pellets were resuspended in 100µl of buffer B + 0.5mM DTT.

2.8.2.1 DNA fragmentation

To fragment the DNA in the optimal length required for ChIP experiments (150-500 bp fragments), samples were incubated with 0.5µl of Micrococcal nuclease (#10011) at 37°C for 20 min, mixing them by inverting the tube every 3 min. This reaction was stopped by adding 10 µl of 0.5M EDTA (final concentration of 50mM) and incubating on ice for 3 min. Nuclei were pelleted by centrifugation at 16.000 x g for 1min at 4°C. The nuclear pellet was resuspended in 100µl of 1X ChIP buffer + PIC and incubated for 10min on ice, inverting the tube every 2 minutes. Samples were briefly sonicated again depending on the type of sample (typically 3 bursts of 20 seconds each, on ice, but this parameter have been modified as commented in 6.6). Lysates were clarified by centrifuging at 10.000 x g for 10min at 4°C. Pellets were stored at -20°C and supernatant (crosslinked chromatin) was kept at -80°C until further use. At the same time, 25µl of the chromatin was extracted from this sample to analyse size and concentration of the DNA fragments (see below).

2.8.2.2 Analysis of the size and concentration of chromatin

125µl of nuclease-free water, 6µl of 5M NaCl (0.2M final concentration) and 2µl of RNase A (10mg/ml) were added to the 25µl of chromatin from 2.8.2.1. Samples were briefly vortexed and incubated at 37°C for 30 min, with mixing by inversion every 5min. Samples were incubated with 2µl of Proteinase K at 65°C for 2h, with constant agitation. DNA was then purified as indicated in X, using 50µl of DNA elution buffer. A 10µl aliquot was run in a 1% agarose gel to observe fragment size, ideally ranging from 150 to 900 bp. DNA concentration of the chromatin sample was determined by diluting DNA in 1/50 (2µl of extracted DNA + 98µl of nuclease-free water) and reading the OD₂₆₀. DNA concentration is OD₂₆₀ x 2500 x 2 (µg/ml).

2.8.3 Chromatin immunoprecipitation

As described in the original protocol, ideally 5 to 10µg of digested, cross-linked chromatin must be used for each immunoprecipitation. However, in this study, all the chromatin sample was used in each immunoprecipitation to maximise the amount of DNA recovered. 75µl of chromatin of each IP were thawed on ice and diluted 1:4 in 1X ChIP buffer + PIC. 10µl of the diluted sample was removed and stored at -20°C (herein referred as 2% input sample). The diluted chromatin was transferred to a 1.5ml Eppendorf tube and as incubated 10µl of α-HA antibody ChIP-grade (Abcam, #ab91110) overnight at 4°C. The next day, the ChIP-grade Protein G magnetic beads were resuspended by vortexing and 50µl of beads were immediately added to the diluted chromatin + HA-antibody and incubated for 4h at 4°C in constant rotation. Magnetic beads were pelleted using a magnetic rack (Cell Signalling Technologies) for 1min. Beads were washed with 1ml low salt buffer (stock was prepared with 300µl of 10X ChIP buffer + 2.7ml of nuclease-free water for each IP), and the beads were incubated at 4°C for 5min in rotation. Magnetic beads were pelleted in the magnetic rack for 1 minute and were washed twice more with low salt buffer. After this, magnetic beads were washed with high salt buffer (100µl of 10X ChIP buffer+ 900µl of nuclease-free ddH₂O + 70µl of NaCl 5M for each IP) at 4°C for 5min with rotation. Magnetic beads were then pelleted in the magnetic rack and supernatant was carefully removed/discarded.

2.8.4 Chromatin elution and reversal crosslink

The 2% input sample was thawed on ice, and 150µl of 1X ChIP elution buffer (75µl of 2X ChIP elution buffer + 75µl of nuclease-free water) was added to each immunoprecipitation and each 2% input sample. Chromatin was then eluted by incubation at 65°C for 30 min with continuous rotation. The supernatant from the beads (which contained the eluted chromatin) was carefully transferred to a new tube, and 6µl of NaCl (0.2M final concentration) and 2µl of Proteinase K (20mg/ml) were added to each sample (including 2% input samples) and incubated 2h to overnight to reversal crosslink.

2.8.5 DNA purification

750µl of DNA binding buffer was added to each sample (this includes each IP and each 2% input sample). 450µl of each sample was transferred to a spin column and centrifuged at 17.000 x g for 30sec. The eluent was removed from the collection tube, and the remaining 450µl was transferred to the column and centrifuged again. After removing the eluant, 750µl of DNA wash buffer was added to each column, which was centrifuged again at 17.000 x g for 30 sec. The eluent was then discarded, and the column was centrifuged at 17.000 x g for 1min. 50µl of DNA elution buffer was added to each column and incubated for 3min at RT. Columns were centrifuged at 17.000 x g for 1 min, and flow-through was passed again through the column to maximise elution of DNA. The concentration of purified DNA was quantified using a Qubit 2 Fluorometric quantitation system (Thermo Fisher) and finally stored at -20°C until used.

2.8.6 ChIP-sequencing

All the procedures described in this section were performed by Glasgow Polyomics (University of Glasgow). DNA libraries were prepared using the NEB Next Ultra II DNA Library Prep Kit (New England BioLabs Inc.) according to the manufacturer's protocol, which included fragments end repair, adaptor ligation, size selection and PCR amplification. The libraries were prepared from 1ng of ChIP DNA with 12 cycles of PCR. Subsequently, the samples were sequenced using an Illumina NextSeq 500 sequencer producing single-end 75 bp long reads, and 20M reads on average. Peak detection, quality trimming and filtering of ChIP-seq reads were performed using sickle and Cutadapt version 1.5 and

mapping of the pre-processed reads to the *Drosophila melanogaster* genome (BDGP6.92). Only the uniquely aligned reads were used for further analysis. The TF binding site enrichment was determined using SICER version 1.1. Window and gap sizes were set at 50 bp and 200 bp, respectively. An FDR cut-off 0.01 was used, and redundant reads were eliminated from analysis. Annotation of ChIP-enriched peaks and identification of enriched motifs was done using HOMER. The transcript database required for chromosome coordinate information was created using the transcriptome annotation 101 files in GFF format (TAIR10).

2.9 Physiology experiments

2.9.1 Lifespan assay

For lifespan assays, adult female flies were kept in standard *Drosophila* medium in groups of 30, transferred every 2 days to fresh vials. During this time, live flies were counted daily until no living flies remained. Knockdown conditions (*CapaR*>*GATAe* RNAi) were compared with both parental controls (*CapaR*/+ and *GATAe* RNAi/+).

2.9.2 Starvation and desiccation assays

Starvation and desiccation assays were performed to assess the tolerance of the flies to stress from different genotypes and adapted from (Terhzaz et al., 2015a). For starvation assays, adult female flies were anaesthetised briefly with CO₂ and kept in groups of 20 in tubes containing 1% aqueous agarose. Vials were kept at all times in an incubator at 22 °C, 55% humidity with a 12:12 light: dark period. During this time live flies were counted every 4h until no living flies could be observed. For desiccation assays, adult female flies were kept in empty tubes and counted every 2h until no living flies could be observed. For all survival assays (lifespan, starvation and desiccation assays), survival data has been expressed as % survival \pm SEM, and all the assays were done in triplicate. Survival data were assessed for significance by the Log-Rank Test using GraphPad Prism 7.0 Software. In all survival assays, mortality was determined by the inability of the flies to resume an upright position after the vial was shaken.

2.9.3 Gravimetric estimations of body water

To measure fly wet body weight, individual adult female flies were anaesthetised on ice and weighed on an AND GR-202 precision balance (analytical weighing to within 0.0001 g). Flies were kept at -80°C for 20 min and subsequently were dried at 60°C for 24h. Dry flies were weighed after reaching RT. The weight of total body water was calculated by subtracting dry weight from wet weight. Water loss over 24h was calculated for each genotype by subtracting water content at 24h from that at 0h. Experiments were run in triplicate with at least 30 flies of each genotype. Data were then processed using GraphPad Prism 7.0 Software and significance was assessed by 1-way ANOVA analysis (Bonferroni Multiple Comparison test) being significant a P-value <0.05 .

2.9.4 Ramsay secretion assay

The Ramsay secretion assay experiments were performed to assess the secretion rates of MTs from flies with different genotypes (Dow et al., 1994, Davies et al., 2019). Petri dishes were prepared by filling them with paraffin wax and depressions were prepared for the secreted medium drops, then, plates were filled with mineral oil. Secreted medium drops contained 50% *Drosophila* Schneider's Medium and 50% of *Drosophila* Saline, stained pink using amaranth dye (see Appendix 9). 9 μl of the medium was pipetted in each well. Intact MTs of adult female flies were dissected in *Drosophila* Schneider's Medium. The MTs were placed in the drops, and one of their ends was wrapped around the pin as shown in Figure 2.9.1. Once the MTs were placed and secreting, the drops were removed from the ureters. The drops secreted by the MT ureters were collected using a fine glass rod exactly every 10min and measured using the microscope graticule. After 30min (3 drop readings), 1 μl of diuretic hormone Kinin (Cambridge Peptides, Birmingham, UK) was added, to the medium drop at a final concentration of 10^{-7}M , and readings were taken for 30 more min every 10min interval. Data were analysed in a Microsoft Excel worksheet (Microsoft Office Package) and posteriorly was processed using GraphPad 7.0 Software. Statistical significance was assessed comparing the relative secretion increase (% Delta) between the samples, ($P<0.05$) and plotted as mean \pm standard error of the mean (SEM).

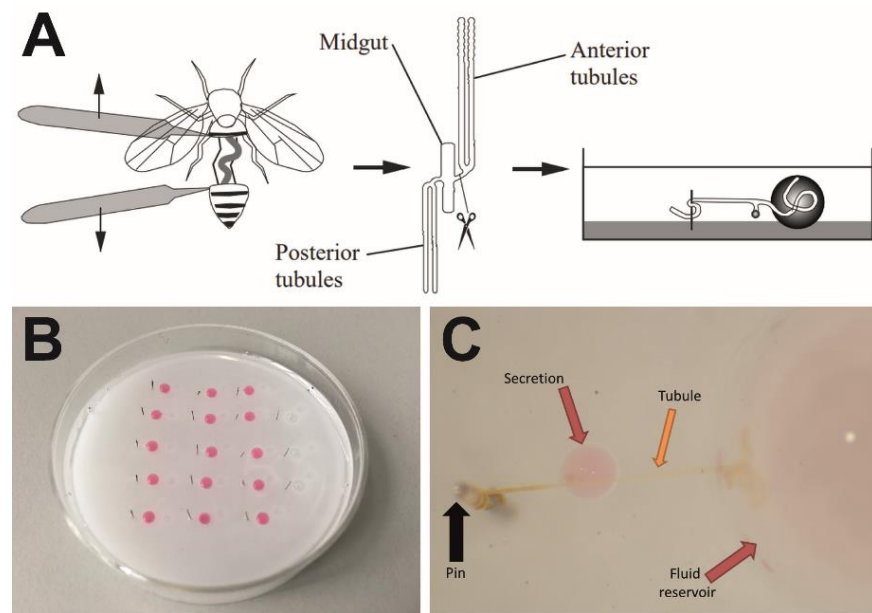


Figure 2.9.1. The Ramsay Secretion Assay.

(A) Illustration of dissection of an adult fly, intact tubules were dissected and carefully placed in the assay plate, with one end immersed in the fluid drop, and the other end surrounding the pin. (B) An experimental assay plate. (C) Magnification of one tubule assay system indicating the essential parts. Adapted from (Davies et al., 2019).

2.10 Bioinformatics

The online databases FlyBase ((Ashburner and Drysdale, 1994)) and FlyAtlas ((Leader et al., 2018, Chintapalli et al., 2007)) were extensively utilised to investigate all candidate genes identified in this project. Identification of putative homologies of genes was performed using the online database Basic Local Alignment Search Tool (BLAST, <https://blast.ncbi.nlm.nih.gov/Blast.cgi> (Boratyn et al., 2013)), by introducing each *Drosophila* gene sequence and searching for its closest homolog in *H. sapiens*. For motif identification in the GATAe sequence, multiple online platforms were employed. Those were: MOTIFSCAN (https://myhits.isb-sib.ch/cgi-bin/motif_scan, (Pagni et al., 2007)), MOTIF (GenomeNet, Japan, <https://www.genome.jp/tools/motif/>, (Kanehisa et al., 2002)), and ScanProsite (<https://prosite.expasy.org/scanprosite>, (De Castro et al., 2006)). The SWISS modelling tool (Waterhouse et al., 2018) was used for generating the 3D structure of the GATAe protein. Gene annotation and gene ontology (GO) analysis of genes were performed using the PANTHER database (Protein ANalysis THrough Evolutionary Relationships, <http://pantherdb.org> (Mi et al., 2016, Mi et al., 2013)) and GoTermMapper (<https://go.princeton.edu/cgi-bin/GOTermMapper>, (Boyle et al., 2004)). Genes with GO terms were classified based on their biological processes and molecular function(s).

Chapter 3 - Genetic Screen of novel Transcription Factors involved in Malpighian Tubule Development and Morphology

3.1 Summary

This chapter is focused on a genetic screen to identify potential TFs involved in MT function. This screen was based on published data describing enriched TF expression levels in the MTs (Wang et al., 2004). As a summary, all TFs in Table 1.2-1 were assessed according to the following selection process: (i) confirmation that those TFs are indeed significantly enriched in the adult and larval MTs, by using the FlyAtlas and FlyAtlas 2 databases. (ii) Identification of those TFs also expressed in the MTs during embryogenesis and which may have a developmental role. (iii) Review of scientific literature for previous studies involving TFs in the MTs and other tissues. (iv) Presence of human TF orthologues, especially in relation to kidney disease. (v) Analysis of loss-of-function via RNAi-based system and morphological phenotyping. As a result of these criteria and data resulting from these analyses, the TF *GATAe* (*CG10278* in Table 1.2-1) was selected as a focus for this research. However, several other TFs have also been identified as potentially interesting to study in the MTs.

3.2 Results

3.2.1 List of most enriched tubule TFs

Among the initial list of more than 30 TFs of Table 1.2-1 (Wang et al., 2004) a relatively small number have been extensively studied and have been shown to be crucial for the MT development. These include *ct* (Hatton-Ellis et al., 2007), *tsh* (Laugier et al., 2005, Denholm et al., 2013), *fkh* (Maruyama et al., 2011), and *pnt* (Kerber et al., 1998). As a result, they have not been considered for further investigation, as this project was to focus on novel functions of TFs.

3.2.2 TFs enrichment in adult MTs

Specific expression of TFs in the adult MTs was confirmed, by interrogating the publicly available resources for gene expression, FlyAtlas (<http://flyatlas.org/atlas.cgi>) and the newly generated FlyAtlas 2 (<http://flyatlas.gla.ac.uk/FlyAtlas2/index.html>, see 1.2.4). Among them, only the TFs that were most enriched in both datasets were considered for further research. Interestingly, one of the most enriched genes in this list (*CG2779*, Table 1.2-1), does not correspond to a TF but encodes for *Mucin 11A* (*Muc11A*) a

protein from the mucin family. *Muc11A* is involved in the formation of the extracellular matrix and exhibits chitin-binding domains (Syed et al., 2008). Curiously, previous research showed that proteins involved in the formation of the extracellular matrix and chitin-related proteins are required for the correct morphology of the MTs (Jack and Myette, 1999). These data suggest that *Muc11A* could have a potential role as well in the development of the MTs. Among all the TFs present in the initial list, a total of 12 were shortlisted as they were highly enriched in the adult MTs but poorly studied to date, in this tissue. These genes were: *GATAe*, *bowl*, *hth*, *STAT92e*, *HNF4*, *ptx1*, *cnc*, *smox*, *hr39*, *bun*, *corto* and *sv*.

3.2.3 TFs expression during embryogenesis

In order to study the specific expression of TFs enriched in the embryonic MTs, different resources were analysed, including the online database BDGP (see 1.1.3). Among the shortlisted TFs, those that are highly expressed in the embryonic MTs, and therefore with potentially important functions in MT development, were considered for further research (Table 3.2-1).

Gene	Expression in embryonic MTs
<i>GATAe</i>	Expressed in all stages
<i>Bowl</i>	Not expressed
<i>hth</i>	Tip cell (Zohar-Stoopel et al., 2014)
<i>STAT92e</i>	Ubiquitous
<i>HNF4</i>	Not expressed
<i>Ptx1</i>	Expressed in MT primordia and tip cell
<i>Cnc</i>	Not expressed
<i>Smox</i>	Faint ubiquitous
<i>Hr39</i>	Not expressed
<i>Corto</i>	Not expressed
<i>bun</i>	Not expressed
<i>sv</i>	Not expressed

Table 3.2-1. Expression of the shortlisted TF in the embryonic MTs, as revealed by BDGP.

According to BDGP in-situ data, the TFs *HNF4* and *sv* are not expressed in the embryonic MTs, but they have been listed due to their importance in the human kidney. On the one hand, the *Drosophila HNF4* closest homologue in humans is *HNF4a*, which is involved in the development of the kidney (Martovetsky et al., 2013). On the other hand, *sv* is the homolog of *pax2*, a well-studied human gene also involved in the development of the kidney (Dressler and Douglass, 1992,

Rothenpieler and Dressler, 1993). The homology details can be found in the following section.

3.2.4 Human TFs orthologues

In this step of the selection process, previous studies and online resources such as Basic Local Alignment Search Tool (BLAST, <https://blast.ncbi.nlm.nih.gov/Blast.cgi>) were curated and analysed to identify human homologues of the selected *Drosophila* TFs and their possible expression in the human kidney. The complete *Drosophila* coding sequence (CDS) of each of the selected genes was examined to find the closest human homolog using BLAST, (summarised in Table 3.2-2). From this, the most interesting TF to be investigated is *GATAe*, which shares high homology with the human GATA factors 4, 5 and 6 (Murakami et al., 2005).

<i>D. melanogaster</i>	<i>H. sapiens</i>	Conserved domains	% Identity of conserved domain
<i>GATAe (isoform A)</i>	GATA-4/5/6	GATA Zinc Finger	52%
<i>bowl</i>	OST2	FOG Zinc Finger	75%
<i>hth</i>	Meis1 and Meis2	Meis_domain, Homeobox domain	56%
<i>STAT92e</i>	STAT5b	STAT_bind, STAT_CCD	41%
<i>Hnf4</i>	HNF4A	HNF4A DBD	67%
<i>Ptx1</i>	PITX2	Homeobox domain	49%
<i>Cnc</i>	NFE2	bZIP_CNC, MDN1 super family	45%
<i>Smox</i>	SMAD3	SMAD2/3 domain, R-SMAD	62%
<i>Hr39</i>	Nuclear receptor Group 5	Lrh-1_like DBD	29%
<i>corto</i>	No conserved domains detected	-	-
<i>bunched</i>	TSC22D1	TSC22 domain	71%
<i>sv</i>	Pax2	PAX domain	85%

Table 3.2-2. *Drosophila* TFs shortlisted and their closest human homologues.

The percentage is displayed according to their similarities in their protein sequences, according to protein-BLAST. Also, the evolutionary conserved domains (comparing *Drosophila melanogaster* and *Homo sapiens*) are included, employing the conserved domains annotation system of BLAST (Marchler-Bauer et al., 2016).

3.2.5 RNA interference lethality screen

To further select which of the 12 TFs from Table 3.2-2 was the most appropriate to be studied in detail, transgenic RNAi lines for each of the 12 TF genes (and *Muc11A*, not included in Table 3.2-3) were obtained from BDSC, (Table 2.1-1) and were individually crossed with the ubiquitous driver *Actin-Gal4*, to identify

possible lethality phenotypes. All crosses were performed at two different temperatures; At 29°C, in which the GAL4/UAS binary system (Duffy, 2002) is optimal with high RNAi expression, and at 25°C, in which the RNAi expression is weaker.

RNAi line	Lethality at 25°C	Lethality at 29°C
<i>GATAe</i>	Pupal lethal	Pupal lethal
<i>Bowl</i>	Not lethal, no detectable defects in MTs	Not lethal, no detectable defects in MTs
<i>Hth</i>	Pupal lethal	Pupal lethal
<i>STAT92e</i>	Not lethal, no detectable defects in MTs.	Not lethal, no detectable defects in MTs
<i>HNF4</i>	Pupal lethal	Pupal lethal
<i>Ptx1</i>	Not lethal, no apparent defects	Not lethal
<i>cnc</i>	Not lethal, no visible defects in tubule	Not lethal, not visible defects in tubule
<i>Smox</i>	Lethal before L3	Lethal before L3
<i>Hr39</i>	Lethal before L3	Lethal before L3
<i>Corto</i>	Not lethal	Not lethal
<i>Bun</i>	Not lethal, no apparent defects in the tubule.	Lethal
<i>Sv</i>	Lethal before the adult stage	Lethal before the adult stage

Table 3.2-3. Lethality experiments.

This table includes information about the phenotypes observed when each TF genes was silenced ubiquitously using *Actin-Gal4*.

3.2.6 RNA interference and specific gene silencing in the MTs

The TFs that induced lethality phenotypes showed in Table 3.2-3 were subsequently crossed with *CapaR-Gal4* and *ctB-Gal4*, two drivers expressed explicitly in the PCs of the MTs (Terhzaz et al., 2012, Sudarsan et al., 2002). These experiments were performed to identify possible morphological defects observed in renal tubules and provided the final evidence supporting the choice of TF for investigation. RNAi lines for *GATAe*, *HNF4*, *smox*, *hr39*, and *sv* genes, which induced premature lethality before the adult stage using *Actin-Gal4*, were crossed with *CapaR-Gal4* and *ctB-Gal4* (Table 3.2-3). Interestingly, at 29°C, only loss of *smox* produced a lethal phenotype during pupal stage using both MT specific Gal4 lines. Also, no evident morphological phenotypes were found in the adult MTs dissected from the other MT-specific knockdown crosses (*HNF4*, *sv*, and *hr39*). However, flies with low levels of *GATAe* using three independent PC-specific Gal4 drivers (*CapaR-Gal4*, *CtB-Gal4* or *c42-Gal4*) developed strong

morphological defects in the MTs (as shown in the following chapters), suggesting a crucial role of this gene in the morphology of this tissue. Based on these data and since *GATAe* has not been studied in the MTs of *Drosophila* although it is one of the most enriched TFs in this tissue (Table 1.2-1), it was selected for further investigation for its novel roles in the MTs.

Furthermore, as mentioned previously in this section, the gene *Muc11A* also appears in the initial list of TFs (Table 1.2-1), even though it is not a TF (Syed et al., 2008). This enrichment in the MTs (Figure 3.2.1A), suggests that *Muc11A* could have a potential role in the development of this organ. Silencing *Muc11A* in the PCs of MTs using *ctB*-Gal4 from stage 9 of embryogenesis resulted in robust morphological abnormalities in the initial segments of adult MTs (Figure 3.2.1B). However, due to time limitations, and that this project was focused on the role of TFs in the renal tubule, it was not possible to study in further detail the role of *Muc11A* in the MTs.

In conclusion, among the shortlisted TF genes, *GATAe* was finally selected to be studied as a novel gene involved in the development of the MTs. Therefore, the involvement of *GATAe* in the development and physiology of the MTs was investigated and results shown in the following chapters of this thesis.

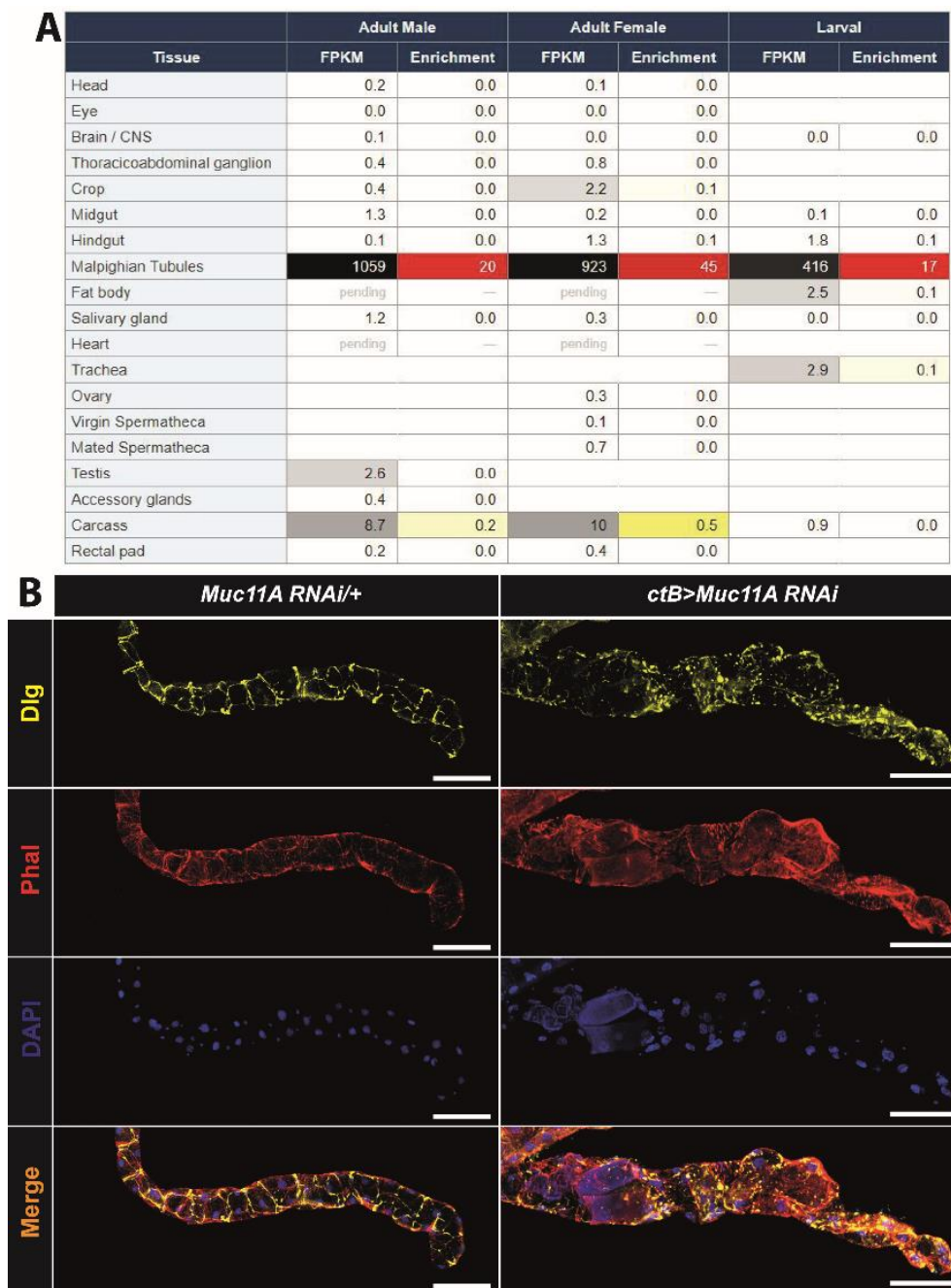


Figure 3.2.1. Low levels of *Muc11A* in the PCs affect the integrity of the adult initial segment. (A) FlyAtlas 2 data show the pattern of expression of *Muc11A*. It is expressed almost exclusively in the larval and adult MTs. (B) Initial segments of control (*Muc11A* RNAi) and *Muc11A* knockdown (*ctB>Muc11A* RNAi) adult MTs, stained with Dlg (yellow), Phalloidin (red) and DAPI (blue). Compared to the controls, *ctB>Muc11A* RNAi MTs exhibit clear morphological abnormalities in their initial segments. Scale bars are 100µm.

3.3 Discussion

Here, a selection process for the assortment of the most appropriate TF to be functionally characterised in MT development was described. This selection process concluded that *GATAe* would be investigated in further detail. It should be noted, however, that some selection criteria could have been improved, as explained below.

First, the absence of gene expression or enrichment of a TF in the embryonic MTs does not necessarily mean that it is not required in this tissue in later stages. For instance, it has been shown that TFs like, *shavenbaby* (*svb*) and *hnt* are expressed in the RNSCs during the adult stage and *svb* maintains this cell population (Bohère et al., 2018, Singh et al., 2007). However, neither of them is present in the initial list of TFs, they are not particularly enriched in the adult MTs according to FlyAtlas. This suggests that there might be other TFs playing significant roles in the MTs, that are either expressed in low levels or at a specific developmental stage. In addition, the data shown in the BDGP database could be occasionally confusing, as *in situ* hybridisation may not detect very low expression of transcript levels and, therefore, not detect possible gene expression in the MTs. For instance, *corto*, a TF that interacts with several Polycomb Group proteins, and has a chromatin structure modulator (Salvaing et al., 2003), is profoundly expressed in the adult SCs (Wang et al., 2004). However, it is not detected in the embryonic MTs according to BDGP, and it is not especially enriched in the adult MTs according to FlyAtlas or FlyAtlas2. Despite this, it would be interesting to investigate if silencing *corto* specifically in the SCs could potentially affect the function of this cell population.

Additionally, the lethality experiments showed in 3.2.6 could present limitations too, as they were performed using RNAi TRiP lines from BDSC. Ideally, these experiments could have also been carried out using different RNAi lines for each gene and then mRNA expression confirmed by qPCR in order to measure the percentage in downregulation of gene expression induced by the RNAi. Accordingly, if the RNAi line utilised did not drive efficient knockdown, it is possible to fail to notice specific phenotypes. Furthermore, lethality experiments are not always entirely reliable, since the complete loss of a particular gene may not induce any morphological phenotype, nor affect fly viability. For example, *ptx1*, a TF highly enriched in the embryonic MTs, would suggest that this gene could play a role in the development of this tissue. However, *ptx1* null mutant flies do not exhibit detectable morphological defects and survive to the adult stage (Vorbrüggen et al., 1997). Still, this gene is required in other tissues in later stages, such as the ISCs in the adult MG (Dutta et al., 2015). Despite the possible troubleshooting of this selection process, the final choice to study *GATAe* is accurate, as shown in the following chapters.

3.3.1 Other TFs to be investigated

Even though the next chapters of this thesis are dedicated to the functional characterisation of *GATAe*, there are several TFs revealed in this study that could be potential candidates for investigation, as they could have novel functions in the MTs, e.g. *Ptx1*. As mentioned previously, *ptx1* is enriched in the MTs in all developmental stages. However, its function in this tissue has not yet been thoroughly studied. Another candidate is *HNF4*, which in *Drosophila* has been characterised for its involvement in lipid mobilisation and insulin signalling response (Palu and Thummel, 2016). In humans, its closest homologue, *HNF4A*, is expressed in the metanephric tubules of the developing kidney (Duncan et al., 1994, Taraviras et al., 1994). Investigating the function of *HNF4* in *Drosophila* MTs could uncover possible homologous functions of this gene between flies and humans.

Also, other TFs in the initial list (Table 1.2-1) have been previously examined in at least one cell type in the MTs. These include *STAT92e* which is expressed in the RNSCs and plays a crucial function in the self-renewal of this cell population (Singh et al., 2007), or *hth*, which has been well characterised in the development of the embryonic MTs (Zohar-Stoopel et al., 2014). However, *hth* is also strongly enriched in larval and adult stages in the MTs, suggesting additional roles in later developmental stages.

3.3.1.1 Mucin 11A

Finally, the possible role of the protein of the mucin family, *Muc11A*, in the morphology of the MTs has also been presented here. Although it has not been possible (due to time limitations) to study the role of mucins in the renal tubule, studying this family of proteins would be interesting for several reasons. First, there are 7 different mucin genes significantly enriched in the MTs in the L3 stage, including *Muc11A* (Syed et al., 2008). In addition, mucins are present in most organisms in the animal kingdom, including humans. In humans, mucins are expressed in various tissues, especially in mucus-producing organs, and are involved in morphogenetic events such as the development of the lungs (Jonckheere et al., 2011), or the intestine (Chang et al., 1994, Bartman et al., 1998), among others. More recent studies have highlighted the importance of

mucins in the progression of different tumours and identified them as potential therapeutic targets for these diseases (Dhanisha et al., 2018, Danese et al., 2018).

Interestingly, a significant proportion of human mucins are directly or indirectly regulated by GATA factors, (Ellis et al., 2013, Jonckheere et al., 2011, Ren et al., 2004). As a big part of this thesis focused almost exclusively on a *Drosophila* GATA factor (*GATAe*), it would be compelling to investigate if this relation between GATA factors and mucins also occurs in *Drosophila*. Thus, these and previous data suggest a role of mucins in the development of the *Drosophila* MTs, and future work should be done to investigate the precise requirement of these proteins.

In conclusion, the research performed in this chapter of this thesis opens a new window for the study of potentially novel functions of genes in the development of the MTs, and further research should be performed in this direction to determine if those genes have indeed a role in this tissue.

Chapter 4 - Analysis of *GATAe* in the Principal Cells of the MTs

4.1 Introduction

This section is focused on the involvement of *GATAe* in the primary and most abundant cell type in the MTs: The PCs. Reduced levels of *GATAe* in the PCs do not produce any detectable defect in the developing (embryonic) MTs. However, the results presented in this chapter demonstrate that *GATAe* performs a vital role in the MTs during metamorphosis and is required for maintenance of MTs in the adult. Silencing *GATAe* using any PC-specific driver induces strong morphological defects in the adult MTs, including a reduction in total cell number, but an abnormal increase of potential RNSCs. In consequence, this results in an impairment in water homeostasis, reduced tolerance to desiccation and starvation stress, and a shorter lifespan.

4.2 Morphological characterisation

4.2.1 *GATAe* is expressed in all MT cell types

As mentioned in the introduction, adult MTs are composed at least by three cell-types: PCs, SCs and RNSCs. In order to investigate which cell types *GATAe* is expressed, two previously generated *GATAe*-Gal4 lines were employed in this study (see Table 2.1-1) (Zhai et al., 2018, Kvon et al., 2014). They contain ~2kb fragments (2267bp for *VT042357* and 2104bp for *VT042358*, that correspond to putative enhancer regions of *GATAe* (Figure 4.2.1) cloned into the pBPGUw vector, which include a basal synthetic core promoter (DSCP, it contains the TATA, Inr and DPE sequence motifs) followed by a GAL4 driver (Pfeiffer et al., 2010, Kvon et al., 2014). Upon combining them with cytosolic GFP, both *GATAe*-Gal4 lines resulted in the same pattern of expression in the MTs, described below. Therefore, herein they will be referred to as *GATAe*-Gal4, or *GATAe*>GFP, when in conjunction with GFP.

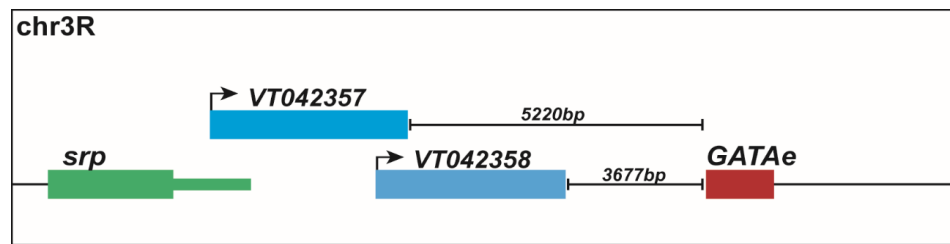


Figure 4.2.1. Schematic representation of the genomic map of *GATAe*-Gal4 lines.

This schematic genomic map represents the genomic location in the chromosome 3R (chr3R) and shows the regions used to generate both *GATAe*-Gal4 lines. The region VT042357 (corresponding to the VDRC line #209818) is 5220bp upstream of the first exon of *GATAe* (red box). It also partially overlaps with the last exon of *srp* (green box). The construct VT042358 (corresponding to the other VDRC line, #205732) is 3677bp upstream of the first exon of *GATAe*. However, it does not overlap with *srp*.

ICC of *GATAe*>*GFP* MTs revealed that at wandering L3 stage (WL3), *GATAe* is expressed in all PCs, but it is not present in SCs (Figure 4.2.2A). In contrast, in the adult stage, *GATAe* is expressed in PCs (Figure 4.2.2B and C), RNSCs (Figure 4.2.2C) and SCs (Figure 4.2.2B). These data also confirmed the pattern of expression of *GATAe* in the adult MG (data not shown), which has been previously reported (Zhai et al., 2018). Also, when *GATAe*-Gal4 was crossed with an RNAi line for *GATAe* (both independent *GATAe* RNAi lines from VDRC and BDSC) at 29°C, it induced 100% lethality shortly after embryogenesis, in accordance to previous reports (Okumura et al., 2005, Okumura et al., 2016). Further qPCR experiments indicated that *GATAe*>*GATAe* RNAi embryos exhibited slight but significantly lower (around 60%) expression levels of *GATAe* mRNA compared to the controls (Figure 4.2.2D). These results confirm the FlyAtlas data and indicate that *GATAe* is present in all three cell types of the MTs in the adult stage, and at least in PCs in earlier stages.

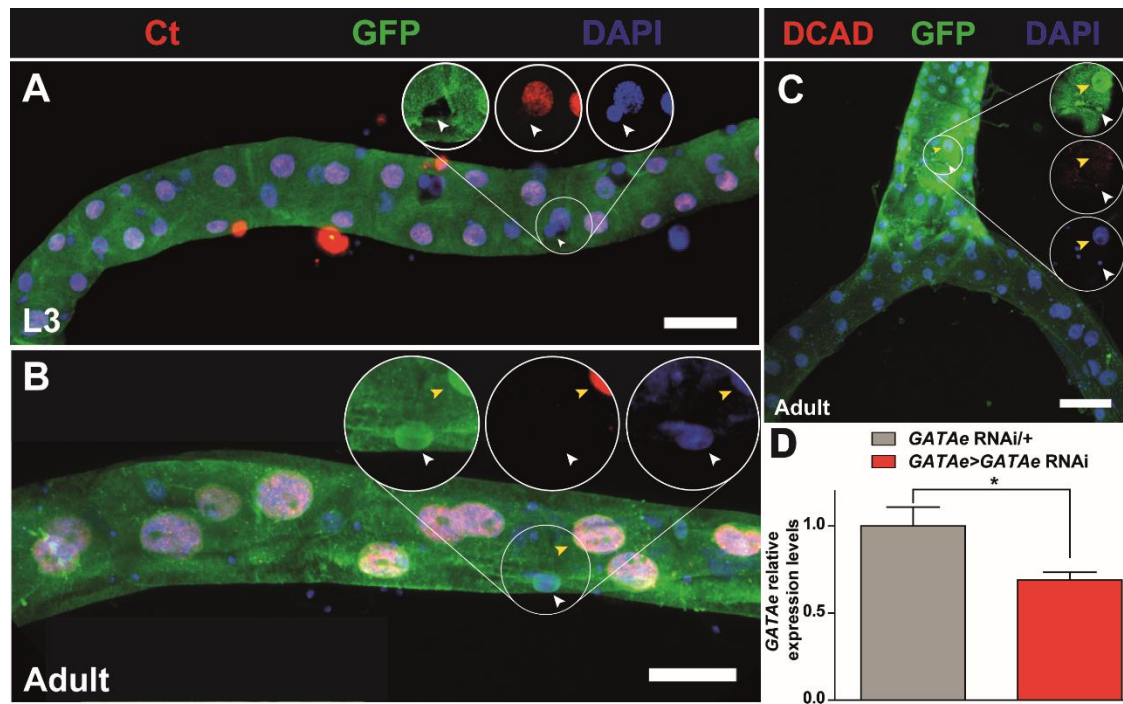


Figure 4.2.2. *GATAe>GFP* reveals the pattern of expression of *GATAe*.

Immunocytochemistry experiments of *GATAe>GFP* indicate the pattern of expression of *GATAe*. (A) Wandering L3 and adult (B) stained with ct (red), GFP (green) and DAPI (blue). *GATAe* is limited to the PCs of the MTs, but not the SCs in WL3 (white arrow in A). However, in the adult stage *GATAe* is expressed in both PCs (yellow arrow in B) and SCs (white arrow in B). (C) Adult *GATAe>GFP* MTs stained with DCAD (red), GFP (green), and DAPI (blue). It shows that *GATAe* is also expressed in RNSCs (white arrow) and PCs of the ureters (yellow arrow). Confocal sensitivity detectors were adapted to obtain a clear image of the GFP in the ureter. However, there is still significant GFP signal in the main segment of this tubule, although even if in this picture is not shown. (D) qPCR experiments show that *GATAe>GATAe* RNAi embryos exhibit reduced levels of *GATAe* mRNA expression compared to the controls. The bar with * indicates significant difference ($p=0.0481$, Student t-test, two-tailed). Scale bars are 50 μ m.

Surprisingly, when *GATAe>GATAe* RNAi embryos were allowed to develop at 26° (where the activity of the GAL4/UAS system is lower as compared to 29°C), they survived until later stages but died during metamorphosis. In addition, *GATAe>GATAe* RNAi pupae exhibited robustly smaller size compared to the controls (Figure 4.2.3). However, further experimentation needs to be done to verify the cause of this reduction in size of *GATAe>GATAe* RNAi pupae.

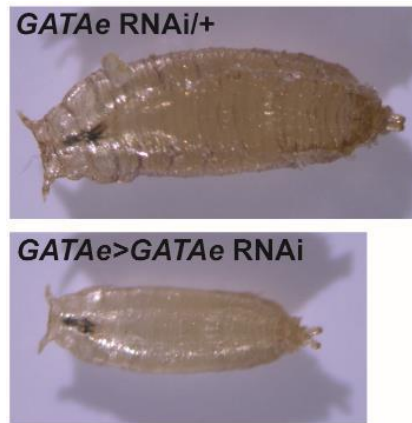


Figure 4.2.3. *GATAe>GATAe* RNAi pupae are smaller compared to the controls.

Compared to the controls (top), *GATAe>GATAe* RNAi pupae, raised at 26°C (bottom) show a reduced size. In addition, *GATAe>GATAe* RNAi die at this stage and never reach adulthood. Both images are on the same scale.

Furthermore, *GATAe>GATAe* RNAi flies were also raised at 18°C. Under these conditions, *GATAe>GATAe* RNAi flies overcame the lethality effects observed at higher temperatures, resulting in the presence of adult flies. However, *GATAe>GATAe* RNAi adult flies died in approximately 7 days after eclosion. Adult MGs and MTs were immunostained and observed at 3 days after eclosion. As expected, adult MGs of *GATAe>GATAe* RNAi flies were extremely short compared to the controls, presented a high presence of trachea, and exhibited extremely altered morphology (Figure 4.2.4B). *GATAe>GATAe* RNAi MTs also displayed morphological defects, similar to the phenotypes obtained driving *GATAe* with PC-specific drivers (Figure 4.2.4D, and see next section).

These strong phenotypes correlate with the requirements of *GATAe* not only in the adult MG (Zhai et al., 2018, Zhai et al., 2017, Okumura et al., 2016), but also in the embryonic development of this tissue (Okumura et al., 2005).

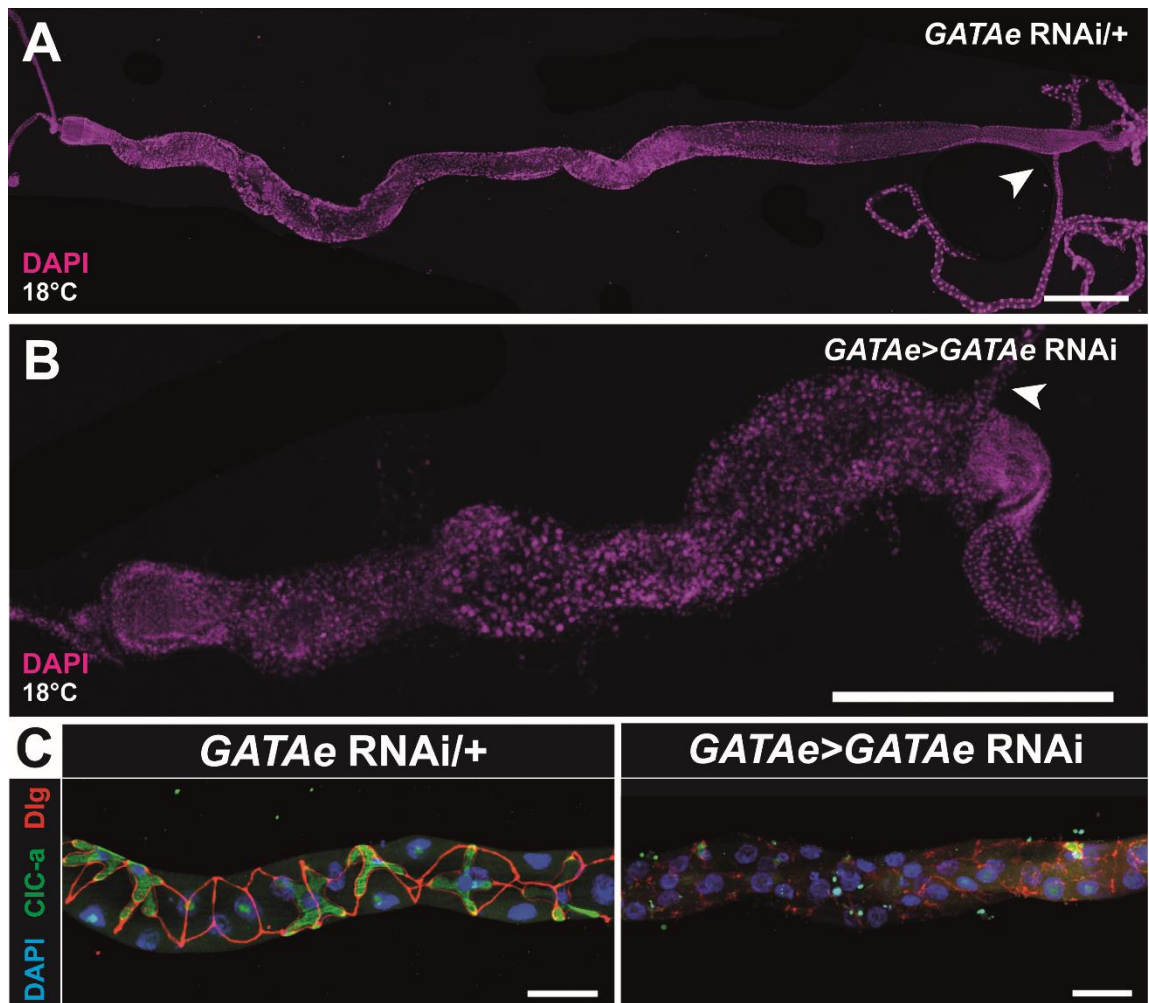


Figure 4.2.4. ICC of adult *GATAe>GATAe RNAi* adult MGs

Control (A) and *GATAe>GATAe RNAi* (B) MGs 3 day-old adult female flies stained with DAPI. Arrowheads indicate the ureters. (C) Adult *GATAe/+* (left) and *GATAe>GATAe RNAi* (right) MTs stained for Dlg (red), Clc-a (green) and DAPI (blue). Scale bars are in A and B, and 500µm and 50µm in C.

4.2.2 Reduced levels of *GATAe* lead to morphological abnormalities of the MTs

Adult *GATAe* knockdown MTs (*CapaR>GATAe RNAi*) exhibited severe morphological deficiencies. First, the length of the tubules is strongly affected, as they are almost four times shorter compared with the controls (775µm in *CapaR>GATAe RNAi* vs 2994µm and 2714µm in the controls, Figure 4.2.5A). Also, their diameter is abnormally increased. Second, *CapaR>GATAe RNAi* MTs display inappropriate organisation of the different cell types that compose the tubule (Figure 4.2.5A and Figure 4.2.7F-J). This strong phenotype in *GATAe* knockdown tubule has been confirmed using two alternative RNAi lines (BL33748, and V10418), and two additional PC-specific Gal4 lines (*ctB-Gal4*, *c42-Gal4*) which phenocopied the morphological defects observed in Figure 4.2.7, (Figure 4.2.5, data not shown).

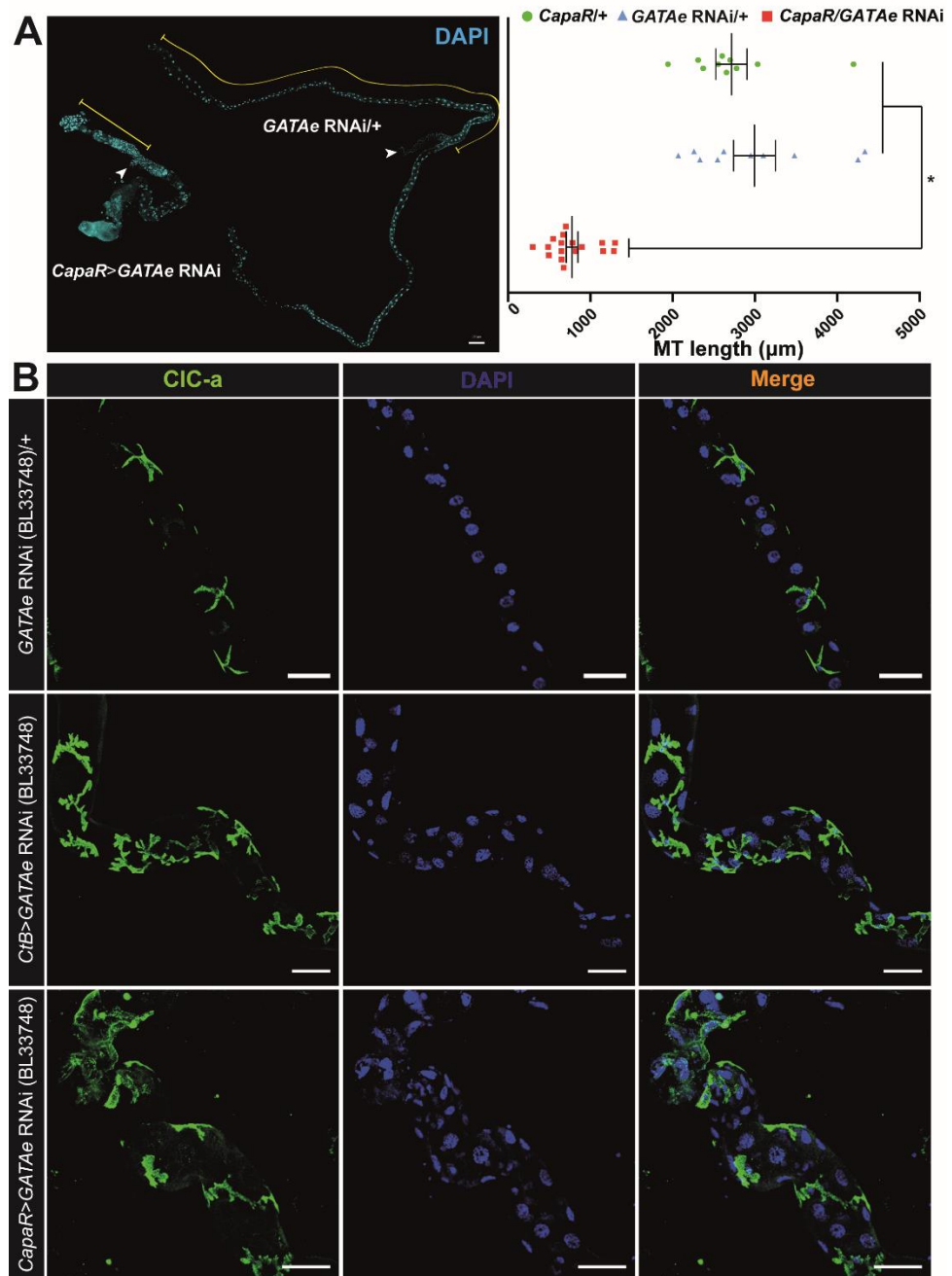


Figure 4.2.5. Length comparison of *GATAe* knockdown MTs.

(A, Left). Control and a *CapaR>GATAe RNAi* anterior MT, stained with the nuclear marker DAPI (Cyan). The length of each tubule is highlighted with a yellow line, and ureters are marked with a white arrow. (Right) Quantifications of MT length of controls and *CapaR>GATAe RNAi* adult MTs. Means are: *CapaR>GATAe RNAi* = $777.5 \mu\text{m} \pm 70.25 \text{ SEM}$, $n=17$, *GATAe RNAi/+* $2994 \mu\text{m} \pm 253.8$, $n=10$, *CapaR/+* $2714 \mu\text{m} \pm 189.2 \text{ SEM}$, $n=10$. $P<0.05$, Student T-test, two-tailed. (B) Adult tubule main segments stained with CIC-a (green, stellate cell marker) and DAPI (blue), using an alternative RNAi line (BL33748). The SC shape appears abnormal compared to controls (top panel), and the morphological defects are similar to the other RNAi lines. This RNAi has been driven with two PC-specific Gal4 lines: *CtB-Gal4* (A-C) and *CapaR-Gal4* (D-F). Scale bars are $100\mu\text{m}$ in A and $50\mu\text{m}$ in B.

Cell number quantifications were performed in order to investigate if *GATAe* knockdown in the PCs induced a reduction of total cell number. Due to the altered morphology of both *CapaR>GATAe RNAi* and *ctB>GATAe RNAi* (BL33748) MTs, it has been difficult to differentiate between SCs and PCs. In consequence,

all the cells present in these MTs have been included in these quantifications (presumably SCs and PCs). Cells which were morphologically similar to RNSCs (described in 4.2.6 and (Singh et al., 2007)) were not included in cell counts. Cell number quantifications indicated that both *CapaR*>*GATAe* RNAi MTs and *ctB*>*GATAe* RNAi (BL33748) adult MTs exhibited a strongly reduced total number of cells compared to all the controls (48 and 65 PCs+SCs in *GATAe* knockdown conditions vs 123, 130 and 128 PCs+SCs in all control conditions, Figure 4.2.6).

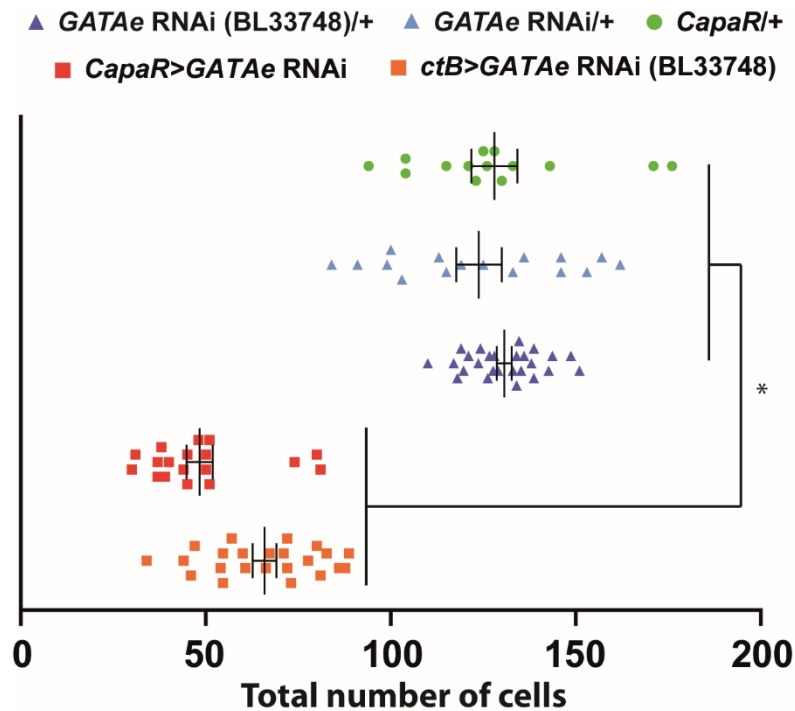


Figure 4.2.6. MTs with reduced levels of *GATAe* show reduced cell number

Cell number quantifications of both PCs and SCs in different control conditions (*GATAe* RNAi/+, *GATAe* RNAi (BL33748)/+ and *CapaR*/+) and PC-specific knockdown conditions (*CapaR*>*GATAe* RNAi and *ctB*>*GATAe* RNAi (BL33748)). *GATAe* downregulation in PCs results in a reduced total number of cells using both independent RNAi lines, compared to their respective controls. Both anterior and posterior MTs were included in the countings. Means are: *GATAe* RNAi/+ = 123.9 ± 6.166 SEM, $n=18$, *GATAe* RNAi (BL33748)/+ = 130.7 ± 1.998 SEM, $n=26$, *CapaR*/+ = 128.1 ± 6.189 SEM, $n=16$, *CapaR*>*GATAe* RNAi = 48.39 ± 3.591 SEM, $n=18$, *ctB*>*GATAe* RNAi (BL33748) = 65.97 ± 3.2 , $n=23$. $P>0.05$, two-tailed Student T-test.

Next, ICC experiments were performed in *CapaR*>*GATAe* RNAi adult MTs using specific cell markers. SCs were stained with CIC-a antibody, which detects explicitly SCs (Cabrero et al., 2014), and results show that the SCs have lost their characteristic stellate cell shape and appear to show a more ramified shape compared to the controls (Figure 4.2.7G'). Moreover, adult MT cell organisation was also affected as in wild-types, they are a single-cell layer tissue composed of two cells in their circumference (Beyenbach et al., 2010). In contrast, *CapaR*>*GATAe* RNAi MTs exhibited an alteration in their number and

organisation of cells, which indicates defects in cell rearrangement (Figure 4.2.7D which presents the typical two-cell organisation in the diameter of the tubule, and Figure 4.2.7I' which has more than two cells in its diameter). PC membrane organisation was also compromised, as shown with the membrane marker Discs large (Dlg), a crucial component of the septate junctions (Woods and Bryant, 1991) (Figure 4.2.11A). However, in *CapaR>GATAe* RNAi MTs, Dlg protein is not localised correctly to the cell membrane suggesting that *CapaR>GATAe* RNAi MTs show defects in cell polarity (Figure 4.2.11F). Furthermore, phalloidin staining, which marks actin cytoskeleton, also indicates that cell shape and polarity is altered in *CapaR>GATAe* RNAi MTs, as, compared to the controls, with visible actin filament defects (Figure 4.2.7H). In summary, this data revealed that *GATAe* expression is crucial in the PCs to ensure appropriate morphology of the adult MTs.

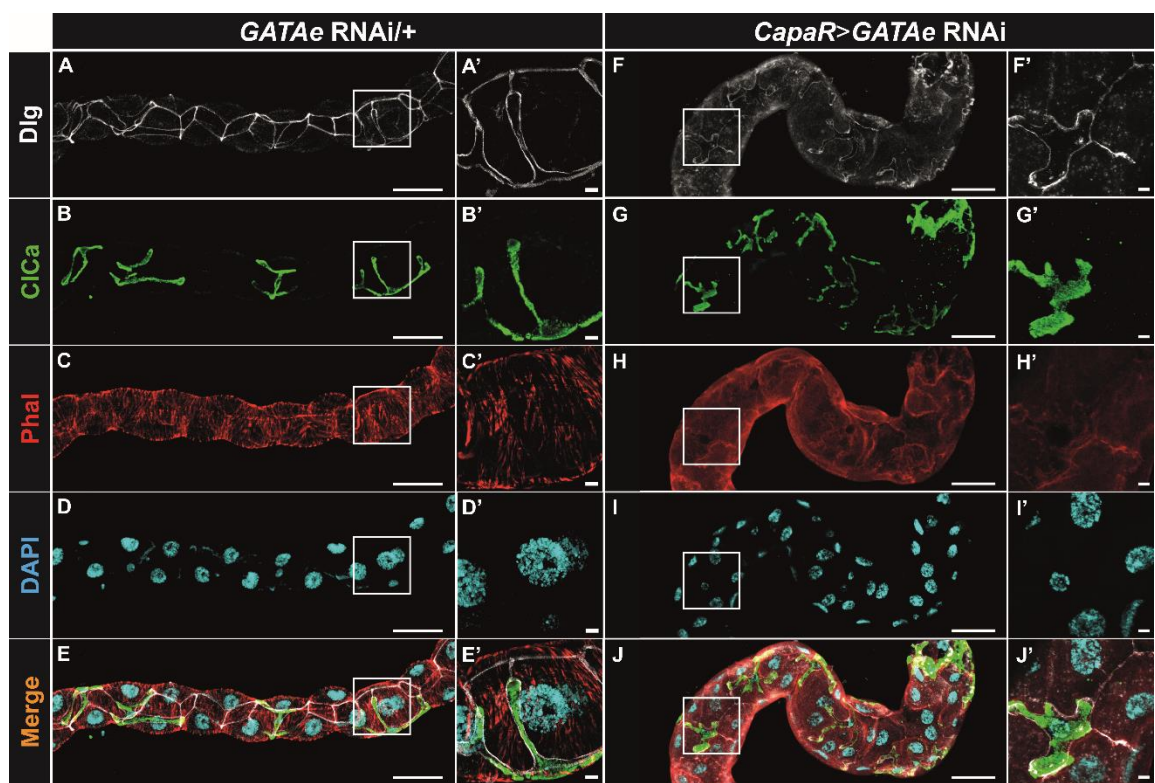


Figure 4.2.7. Reduced levels of *GATAe* in PCs induces strong morphological defects.

ICC with which antibody of control (A-E') and *GATAe* knockdown using *CapaR*-Gal4 (F-J'). Merged images show the overlay of Discs Large (Dlg, staining the septate junctions, white), ClCa (For SCs, green), Phalloidin (F-actin, red) and DAPI (Nuclei, Cyan). (A'-J') Close-up images are showing the shape of an SC, in the areas highlighted in white boxes in A-J, respectively. Scale bars are 50µm (A-J) and 5µm (A'-J').

It has been previously shown that alterations in tumour suppressor genes induce similar phenotypes as the ones caused by knocking down *GATAe* in the PCs (Zeng et al., 2010). One of the most striking phenotypes of hyperplasia in the MTs was

observed upon activation of Ras signalling in the RNSCs, which transformed normal RNSCs into malignant cancer stem cells (Zeng et al., 2010). Therefore, Ras signalling was constitutively activated (Wu et al., 2010, Zeng et al., 2010) in tubule PCs to investigate if it also induced tumorigenesis. Interestingly, *CapaR>Ras^{V12}* MTs deficiencies show striking morphological similarities to *CapaR>GATAe* RNAi MTs; they also show irregularities in cell organisation compared to the controls (Figure 4.2.8). However, *CapaR>Ras^{V12}* MTs seem to exhibit a milder phenotype compared to *CapaR>GATAe* RNAi MTs, as their length does not seem to be as affected as in *CapaR>GATAe* RNAi MTs. These results suggest that *GATAe* could potentially be related to Ras signalling in the MTs. In addition, Ras signalling was silenced in *CapaR>GATAe* RNAi MTs through the induction of an RNAi for *Ras85D*, to investigate if low levels of Ras signalling could counter the morphological phenotypes obtained when *GATAe* was silenced. Additionally, rescue experiments have been performed, by reducing both *GATAe* and Ras signalling in the PCs (combining a *GATAe* RNAi construct and the dominant negative form of *Ras85D*, *Ras^{N17}*). However, adult *CapaR>GATAe* RNAi *Ras^{N17}* MTs completely phenocopied the morphological anomalies of *CapaR>GATAe* RNAi MTs (Figure 4.2.8B). These results suggest that the phenotype observed upon *GATAe* downregulation in PCs could be caused by the activation of additional pathways as well as Ras signalling. Another interesting possibility is that the altered levels of *ras85D* observed upon *GATAe* RNAi in the PCs (as showed in later sections, see Figure 4.2.16) come from the RNSCs. Previous data support this hypothesis, and showed that Ras signalling is overexpressed in RNSCs when they have transformed to “cancer” stem cells (Zeng et al., 2010). This would potentially explain why silencing Ras signaling in PCs does not compensate the morphological defects caused by loss of *GATAe*.

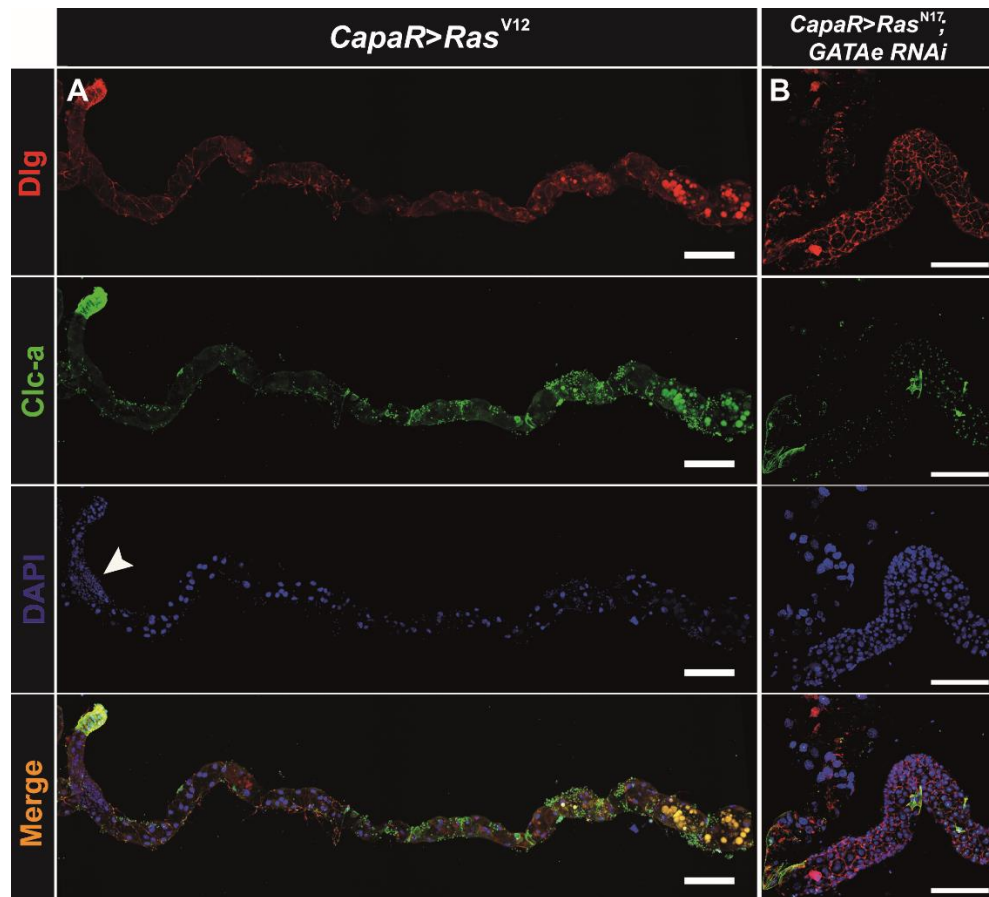


Figure 4.2.8. Activation of Ras signalling in PCs induces morphological defects.

(A) ICC experiments of *CapaR>Ras^{V12}* adult MTs, marked with Dlg (red), CIC-a (green), and DAPI (blue). They show morphological similarities with *CapaR>GATAe RNAi* MTs, shown in Figure 4.2.7. There is also a high increase of small nuclei cells in the ureters (white arrowhead). (B) Expression of a dominant negative form of *Ras85D* (*Ras^{N17}*) does not rescue the morphological defects caused by *GATAe* knockdown. Scale bars are 200µm in A and 100µm in B.

CapaR-Gal4 is expressed in additional tissues than the MTs (e.g. gut musculature, Terhzaz et al., unpublished). Given that *GATAe* is expressed in the MG during all stages of development and in the adult stage (Okumura et al., 2005, Okumura et al., 2016), it is possible that the defects observed using this line could be a consequence of *GATAe* downregulation in the MG. To rule out this possibility, MGs of *CapaR>GATAe RNAi* flies were immunostained and observed. No detectable differences were detected in *CapaR>GATAe RNAi* MGs when compared to the controls (Figure 4.2.9). Their diameter was as MGs from control flies, contrasting with the phenotype previously reported when *GATAe* was lost in the ISCs, which results in a reduction in the diameter of the MG (Okumura et al., 2016) (Figure 4.2.9F). Also, the pattern of expression of the marker Hindsight (*Hnt*) was indistinguishable to the controls and similar to other reported publications (Baechler et al., 2016) (Figure 4.2.9H). Altogether, these results indicate that the deficiencies caused by *GATAe* knockdown in the PCs do not affect MG morphology.

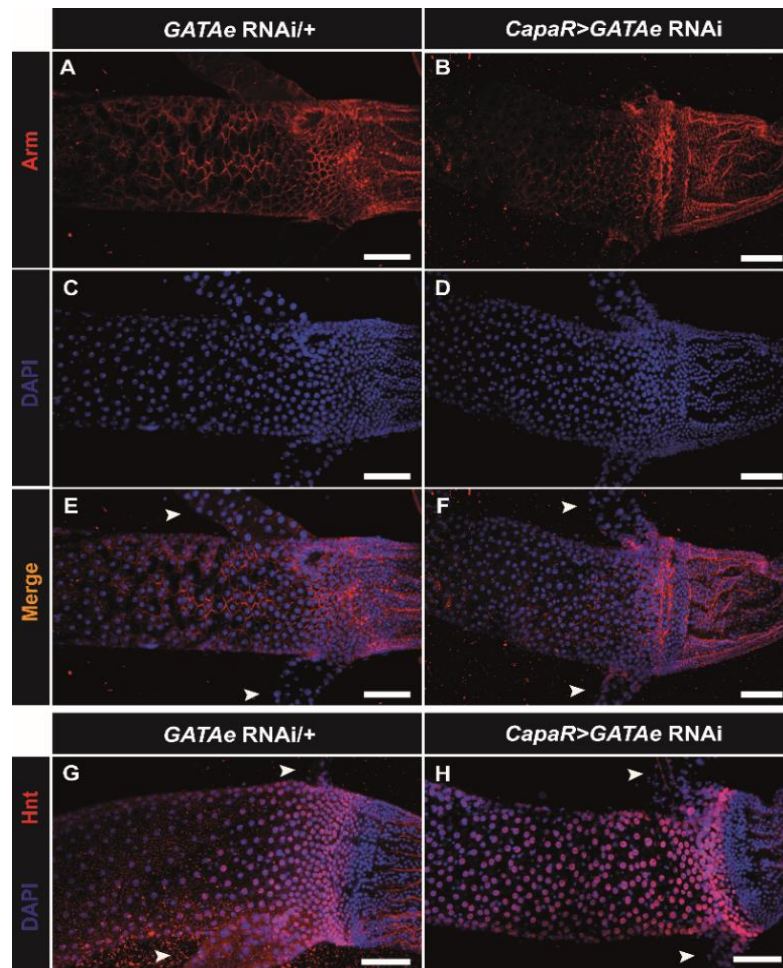


Figure 4.2.9. Silencing *GATAe* with *CapaR-Gal4* does not affect MG morphology.

Maximum projections of adult posterior MGs, including a portion of the hindgut, stained with Arm (red) and DAPI (blue) (A-F) or with Hnt and DAPI (G and H). No significant differences in structure or diameter can be detected in *CapaR>GATAe* RNAi MGs. Also, the organisation of Hnt⁺ cells in *CapaR>GATAe* RNAi MGs are similar to the wild-type.

4.2.3 *GATAe* is not essential during the embryonic development of the MTs

Generally, defects in the morphology of the MTs are due to developmental programming defects during embryonic development, resulting in non-functional MTs and premature lethality (Campbell et al., 2010, Hatton-Ellis et al., 2007). In addition, the MT-specific expression of *GATAe* observed in all embryonic stages (Okumura et al., 2005) suggests that *GATAe* could be involved in MT development. It has been demonstrated that *CapaR* is exclusively expressed in the PCs during all larval stages and in the adult and probably during embryogenesis (Terhzaz et al., 2012). To confirm that *GATAe* is not required during embryogenesis this, *GATAe* was silenced using *ctB-Gal4*, which drives gene expression from stage 9 of embryonic development (Sudarsan et al., 2002). MTs from both control and *ctB>GATAe* RNAi embryos were compared at two different developmental stages; at stage 14, when the MTs are actively

elongating and migrating through the embryo (Figure 4.2.10A and B), and stage 16, when the embryonic development of the MTs is mostly complete (Figure 4.2.10A' and B'). Surprisingly, *ctB>GATAe* RNAi embryos did not show any detectable alterations in size, shape, migration or polarity of their MTs compared to the controls and at both the stages of embryogenesis tested (Figure 4.2.10B and B'). To further confirm these findings, embryonic MTs were also analysed using a null *GATAe* mutant (*GATAe*^{-/-}). This mutant line was generated by a homologous recombination-based gene targeting technique, in which the coding region of *GATAe* was replaced for a copy of the *w*⁺ gene (Okumura et al., 2016). The results obtained using the *GATAe*^{-/-} line were similar to the observations using the RNAi line (Figure 4.2.10C). However, *GATAe*^{-/-} specimens die shortly after embryogenesis, probably due to *GATAe* critical functions in other tissues such as the MG, as reported in previous studies (Okumura et al., 2016, Okumura et al., 2005). Only a small proportion survive until L1 stage and ultimately dies. For this reason, it has been unfortunately impossible to examine any morphological phenotype in *GATAe*^{-/-} adult MTs. Nevertheless, these data indicate that although *GATAe* is expressed in the embryonic MTs, it is not essential for primordia polarity, proliferation or migration. However, it is necessary in later developmental stages for proper MT differentiation.

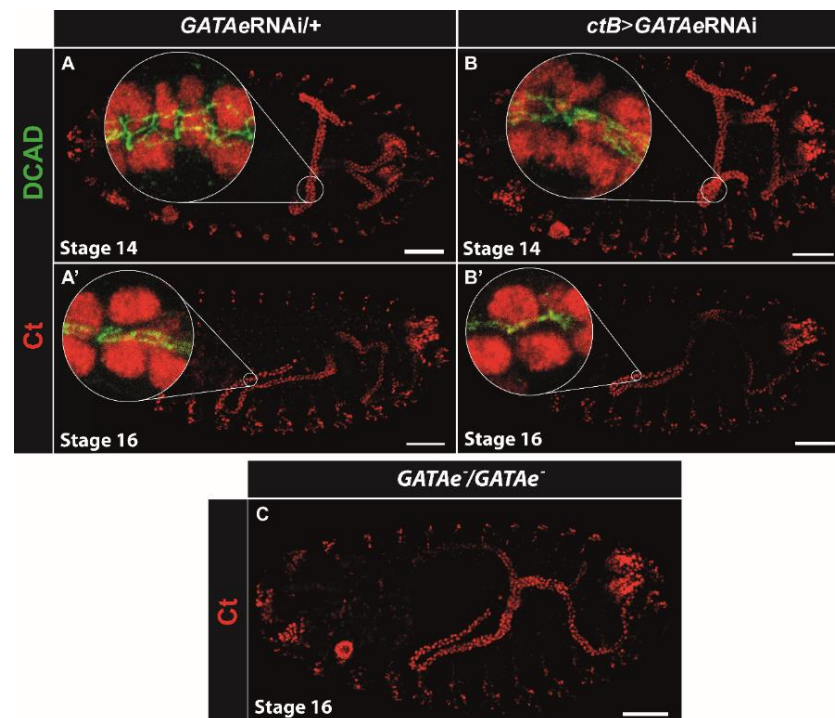


Figure 4.2.10. *GATAe* knockdown embryonic MTs display wild-type morphology.

(A-B') MTs in embryos of Stage 14 and 16 show no detectable differences in cell migration or polarity between *GATAe* knockdown and control embryos. Tubules were stained with ct (red) and DE-Cadherin (DCAD, green) antibodies. Inserts: optical projections of the circled areas, stained

with ct and DCAD, which labels the adherent junctions. (C) Homozygous null mutant for *GATAe* (*GATAe^{-/-}/GATAe^{-/-}*) embryo shows a normal MT morphology. Scale bars are 50µm.

4.2.4 *GATAe* is required during metamorphosis

CapaR>*GATAe* RNAi flies were dissected at different larval and pupal stages to identify the developmental period during which the morphological phenotype occurs. MTs examined at any larval stage (L1, L2 and WL3) show normal characteristics compared to control (Figure 4.2.11 A'-D'). However, at 48h after puparium formation (APF), MTs displayed strong defects which phenocopied the morphological abnormalities showed in the previous section (Figure 4.2.11 E'-H').

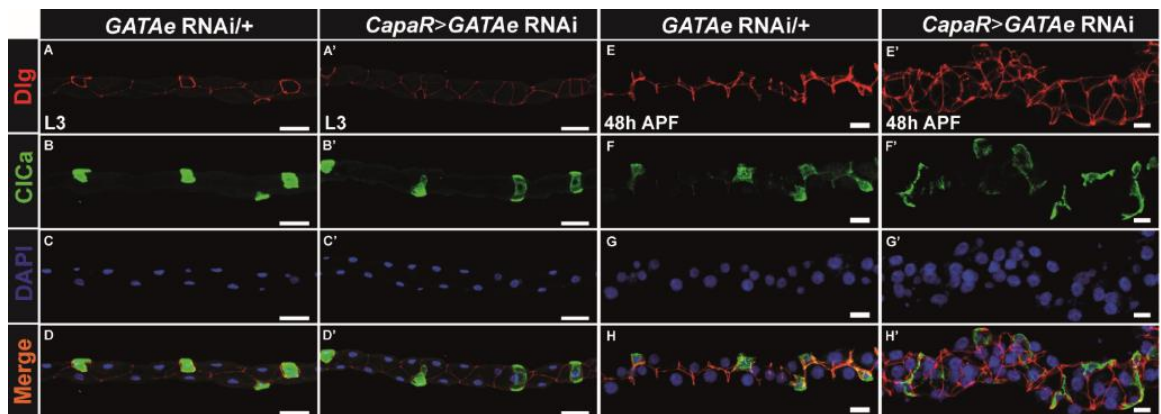


Figure 4.2.11. *CapaR*>*GATAe* RNAi MTs present morphological deficiencies from during metamorphosis.

(A-D') Comparison of third instar larval stage control (A-D) and knockdown tubules (A'-D'). Merge images show an overlay of Dlg (red), ClCa (green) and DAPI (blue). No differences in morphology or polarity were observed. (E-H') Control (E-H) and 48h after pupa formation *GATAe* knockdown tubules (E'-H'). During the pupal stage, *CapaR*>*GATAe* RNAi tubules look disorganised, and the cell arrangement is abnormal (G') Scale bars are 50µm.

Additionally, to precisely pinpoint the developmental stage from which *GATAe* is critically required, the *CapaR*-Gal4;TubulinGal80^{ts} driver (temperature sensitive *CapaR*-Gal4, herein referred as *CapaR^{ts}*, see 1.1.3.3), and *GATAe* RNAi lines were crossed. Parental lines were allowed to lay eggs at either 18°C or 29°C and then switched between these two temperatures at appropriate stages of the development (L1, L2, wandering L3, pupa, or until adult stages). As expected, *CapaR^{ts}*>*GATAe* RNAi adult flies raised at 18°C (where Gal80 inactivates GAL4 activity) showed wild-type MT morphology while adult flies kept at 29°C during all developmental stages (where GAL4 activity is activated) were phenotypically identical to *CapaR*>*GATAe* RNAi MTs (data not shown). Interestingly, *CapaR^{ts}*>*GATAe* RNAi animals kept at 18°C until wandering L3 stage and then

switched at 29°C, manifested morphological defects in their MTs later in the adult stage (Figure 4.2.12B). In contrast, *CapaR^{ts}*>*GATAe* RNAi raised at 29°C until wandering L3 stage and then transferred to 18°C did not show any detectable tubule phenotype compared to control (Figure 4.2.12A). A schematic diagram of the GAL80^{ts} experiments for the crosses performed at different temperatures is shown in Figure 4.2.12C. In conclusion, and together with the data illustrated in Figure 4.2.11, *CapaR^{ts}* experiments show that *GATAe* is required from the L3 stage onwards, probably during early metamorphosis (between 0 and 48h APF). Curiously, during metamorphosis, wildtype MTs shrink half of their length and elongate back to their original size just before adult eclosion (Wessing and Eichelberg, 1979). Therefore, *GATAe* could potentially play a role in this elongation process.

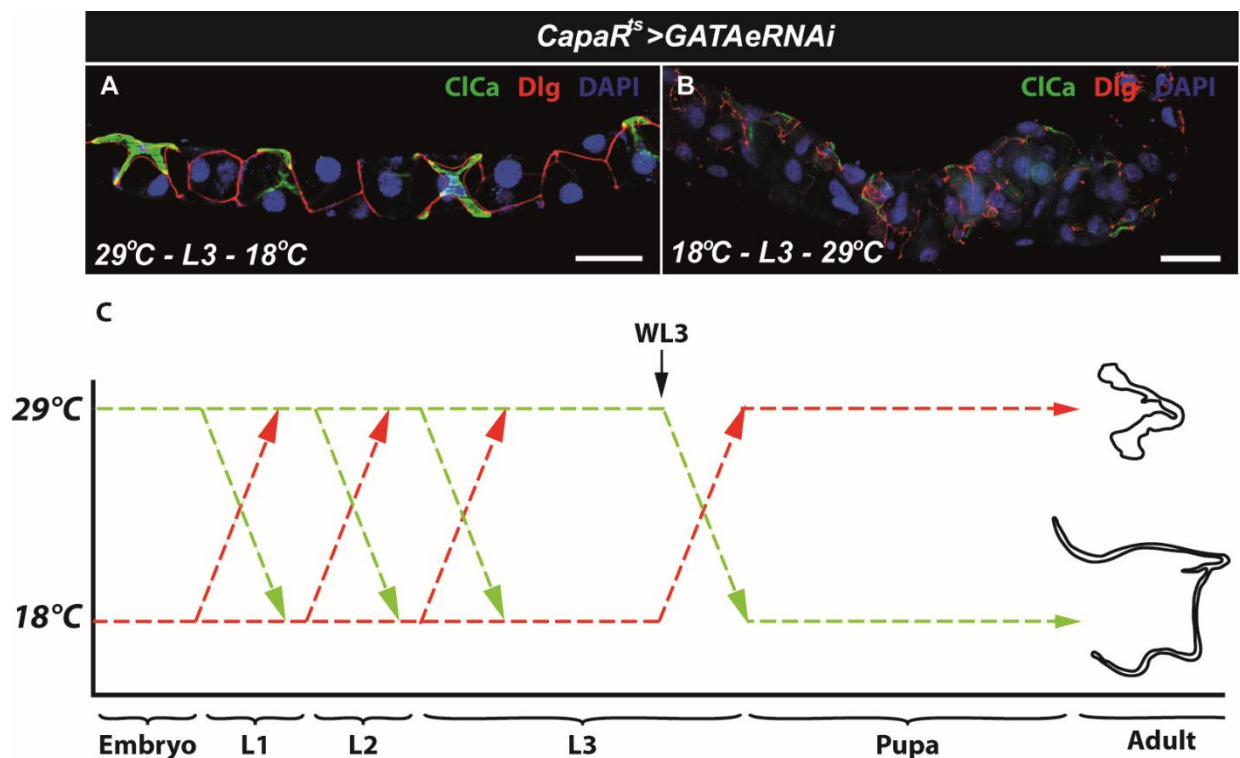


Figure 4.2.12. Developmental defects using *CapaR^{ts}*.

(A) A representative MT main segment of adult *CapaR^{ts}*>*GATAe* RNAi fly that has been raised at 29°C until wandering L3 stage and then switched to 18°C. The MT appears indistinguishable to wild-type. (B) A representative MT main segment of adult *CapaR^{ts}*>*GATAe* RNAi fly that has been raised at 18°C until wandering L3 stage and then switched to 29°C. In this case, the MT shows a strong morphological phenotype. Scale bars are 50µm. (C) *CapaR^{ts}*>*GATAe* RNAi were raised at 18°C and switched to 29°C to activate GAL4 expression and block *GATAe* expression in the PCs during all stages of development until L3 stage (green arrows). In all these cases, the emerging adult flies showed MTs with normal characteristics. In contrast, when these flies were raised at 29°C and switched to 18°C at the same stages until WL3 (red arrows), the MTs from the emerging adult flies exhibited strong morphological defects. L1, L2 and L3 stand for larval stage 1, 2 and 3, respectively. WL3 stands for wandering L3.

4.2.5 Adult MTs maintenance requires the function of *GATAe*

Following the observations that MT function of *GATAe* is needed exclusively from metamorphosis, its possible role in the maintenance of adult morphology and homeostasis was investigated. To perform this, *CapaR^{ts}>GATAe* RNAi flies were raised at 18°C to allow their MTs to develop under normal conditions. After adult eclosion, flies were transferred to 29°C to activate expression of the *GATAe* RNAi construct. Adult flies were aged, and their MTs were collected for immunostaining every 7 days for a total period of 28 days. MTs of flies up to 7 days after RNAi activation showed normal characteristics (regular PC organisation and typical “stellate” shaped SCs) compared to the controls (Figure 4.2.13A). However, MTs started to exhibit morphological defects of cell shape and organisation after 14 days, and after 21 days, morphological abnormalities in the general tubule shape and cell membranes were noticed (Figure 4.2.13A at 28D). In addition, these MTs also exhibited an increased presence of cells positive for Hindsight (*Hnt*⁺), a previously reported marker for RNSCs (Bohère et al., 2018, Baechler et al., 2016), indicating an increase of these type of cells (Figure 4.2.13B) from 14 days.

Altogether, these data suggest that *GATAe* is not only required to maintain the morphology of the MTs during the pupal stage, but it is also actively playing a role in the adult stage to ensure the integrity of this tissue.

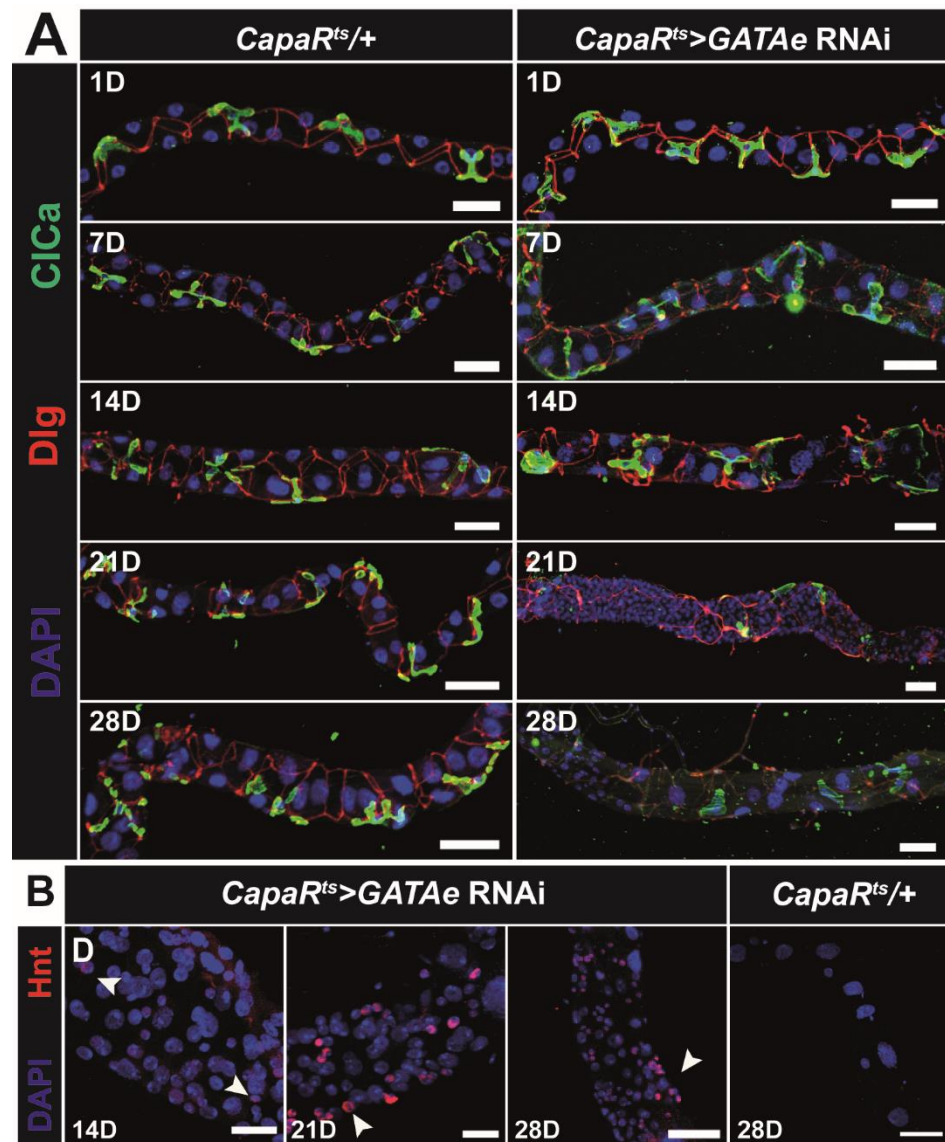


Figure 4.2.13. GATAe is required for adult maintenance of the MTs.

(A) Main segments of adult *CapaR^{ts}>GATAe RNAi* MTs that have been raised at 18°C until eclosion and then switched to 29°C, stained with DAPI (blue), Dlg (red) and CIC-a (green). After 7 days, *CapaR^{ts}>GATAe RNAi* show wild-type characteristics. However, after 14 days they start to gradually exhibit critical morphological defects and cell proliferation. (B) Magnifications of main segments of 14 days, 21 days and 28 day-olds switched MTs, stained with Hnt (red) and DAPI (blue). After 14 days of RNAi induction, there is the presence of Hnt⁺ cells. A control of a *CapaR^{ts}/+* adult main segment 28 days after temperature switching is shown in the right. Scale bars are 50µm, except for B.14D and B.21D that is 20 µm.

4.2.6 Loss of *GATAe* induces cell proliferation

CapaR>GATAe RNAi MTs showed a robust increase in the number of small nuclei cells (potential RNSCs), in all the different regions of the MTs including the ureter and lower tubule (Figure 4.2.14 and Figure 4.2.15 for representative examples). Since the degree of proliferation of these cells was variable between preparations, cell counting experiments have been performed in representative 50µm² of their lower ureter regions. Quantification of cell nuclei showed that *CapaR>GATAe* RNAi MTs showed a significant increase in the number of potential RNSCs in the lower ureter when compared to the controls (Figure 4.2.14H). Those cells were identified as potential stem cells (RNSCs) since they were positive for Armadillo (Arm⁺), the *Drosophila* homolog of β-Catenin, which has been previously reported as a marker for stem cells, present in both RNSCs and ISCs (Ohlstein and Spradling, 2006, Singh et al., 2007).

In adult MTs, RNSCs are limited only in ureters and lower tubules (Singh et al., 2007, Sözen et al., 1997). These over proliferating cells present in *CapaR>GATAe* RNAi MTs also express additional RNSCs markers such as DE-Cadherin (DCAD, Figure 4.2.14, (Li et al., 2015)). Furthermore, they were Hnt⁺ (Figure 4.2.15G). Finally, a large part of these cells, but not all, were found to be positive for Delta (Dl, Figure 4.2.15A), a transmembrane protein from the EGF family, ligand of the Notch signalling pathway (Fehon et al., 1990), and again present in both ISCs and RNSCs (Li et al., 2015, Ohlstein and Spradling, 2007).

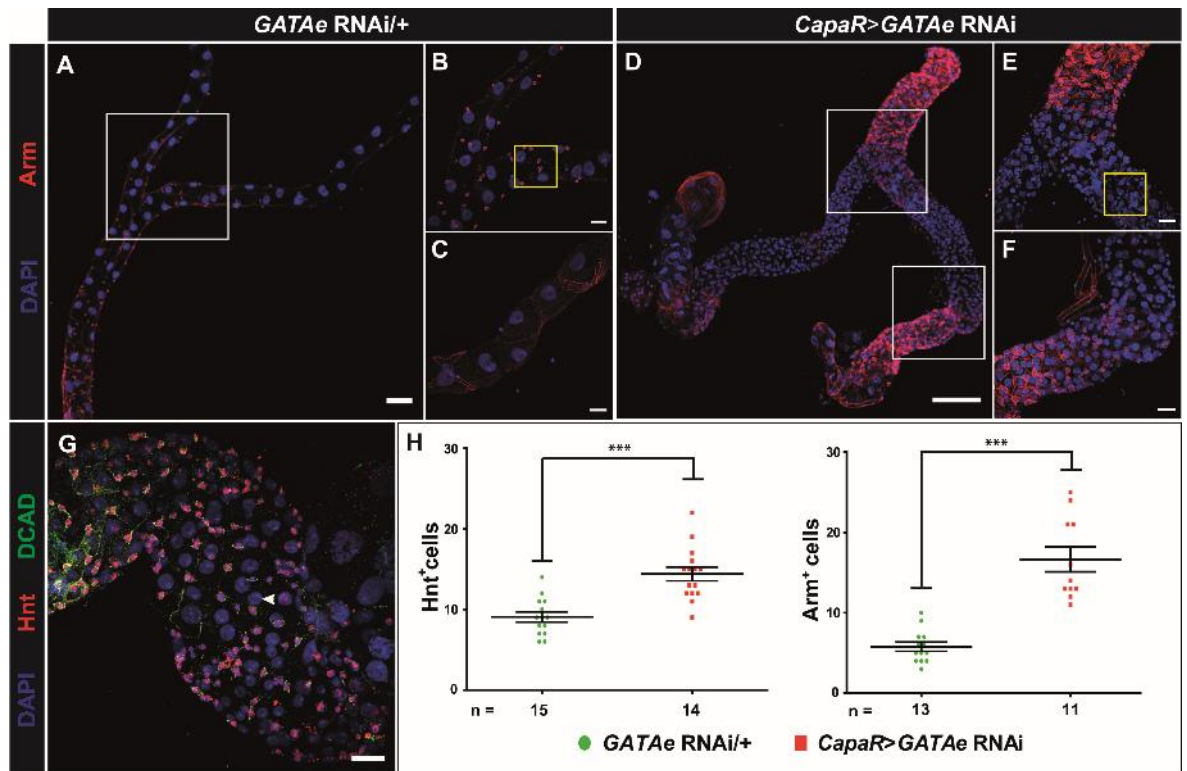


Figure 4.2.14. *GATAe* knockdown in PCs induces proliferation of RNSCs.

Control (A-C) and *CapaR>GATAe* RNAi (D-G) adult MTs. *CapaR>GATAe* RNAi MTs show presence of Arm⁺ (D-F) and Hnt⁺ (G), not only in the ureters but all along the MTs. B, E and F are magnifications of the boxes of A and D, and B is a magnification of a region of the main segment, not present in A. White arrows in A indicate intermediate-sized nuclei that are Arm⁻, DCAD⁻ and Hnt⁻, which could correspond to a cell type previously named Renal blast (Singh et al., 2007) (H) Hnt⁺ (left) and Arm⁺ (right) cell number quantifications of *CapaR>GATAe* RNAi and control 50µm² regions of lower ureters (representative regions are shown in yellow squares of B and E). Scale bars are 50µm for A and D, 20µm for B, C, E, F and G. Lower ureters from *CapaR>GATAe* RNAi exhibit a significant increase of Arm⁺ and Hnt⁺ cell number compared to the controls. Data are presented as mean ± SEM * P < 0.0001, Student t-test, two-tailed. Means for Hnt⁺ cells are: *GATAe* RNAi/+ = 9.071 ± 0.62 SEM, n=14, *CapaR>GATAe* RNAi = 14.4 ± 0.84 SEM, n=15. Means for Arm⁺ are: *GATAe* RNAi/+ = 5.769 ± 0.56 SEM, n=13, *CapaR>GATAe* RNAi = 16.64 ± 1.545 SEM, n=13. *** indicates statistical significance (P<0.001, Student t-test, two-tailed).

In addition to stem cell markers, the mitosis marker Phospho-histone H3 (PH3) (Micchelli and Perrimon, 2006) has been used to determine if these cells were over proliferating. In adult MTs, RNSCs are relatively quiescent and rarely divide, (it has been suggested that their division rates are around 2 RNSCs per every five pairs of MTs, (Zeng et al., 2010, Li et al., 2015)). However, in *CapaR>GATAe* RNAi MTs, several PH3⁺ can be observed in different regions of the MTs (not only in ureters and lower ureters), indicating continuous cell proliferation (Figure 4.2.15B). Furthermore, numerous cells of intermediate nucleus size (between RNSCs and PCs) were detected in *CapaR>GATAe* RNAi MTs (white arrows in Figure 4.2.15 G and Figure 4.2.15A). Given their protein expression profile (Hnt⁻, DCAD) they could correspond to the immature mitosis products of RNSCs, already described in previous publications (Singh et al., 2007). Altogether, these

observations presented indicate that *CapaR>GATAe* RNAi MTs show an overproliferation as well as mislocalisation of RNSCs.

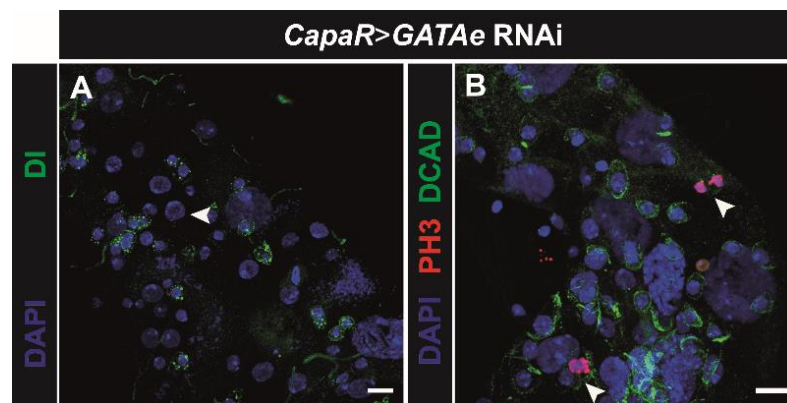


Figure 4.2.15. Alternative RNSC and proliferation markers.

(A) *CapaR>GATAe* RNAi adult MTs stained for DI (green) and DAPI (blue). They show presence of DI⁺ cells, but at lower levels compared to other stem cell markers such as DCAD, Hnt, or Arm. (B) *CapaR>GATAe* RNAi stained with the mitotic marker PH3 (red), DCAD (green) and DAPI (blue). There is also occasional presence of PH3⁺ cells (white arrows), indicating cell proliferation. White arrow in A indicate intermediate-sized nuclei that are DCAD⁻ and DI⁺ probably corresponding to Renal blasts (Singh et al., 2007). Scale bars are and 10µm.

4.2.7 Reduced levels of *GATAe* modulate cancer-related gene expression

Numerous cancers are initiated by a failure in the control of proliferation and differentiation of stem cell populations, including those of the *Drosophila* MT (Reya et al., 2001, Li et al., 2015, Zeng et al., 2010, Parvy et al., 2018). Since *GATAe* downregulation in the PCs induces cell proliferation and abnormal growth in the adult MTs, expression levels of different key genes that regulate these processes have been measured. *Ras Oncogene at 85D* (*Ras85D*) and the tumour suppressor *warts* (*wts*) (Ren et al., 2010) genes expression have been measured to determine the genetic pathways involved in this phenotype; as well as the expression of downstream apoptosis-related effectors, e.g., *death-associated Inhibitor of Apoptosis 1* (*DIAP1*), *buffy*, and *death executioner bcl2* (*debcl*) (Terhzaz et al., 2010).

CapaR>GATAe RNAi MTs exhibit a significant upregulation of *DIAP1* and *Ras85D* gene expression (2.5 and 3-fold respectively, Figure 2.2-10), with downregulation of *debcl* and *wts* (0.7 and 0.6-fold respectively); while no significant difference in *buffy* expression levels was observed in *GATAe* knockdown MTs. These MTs also exhibited a strong upregulation of

Adenosylhomocysteinase like 1 (Ahcyl1), (~5-fold), a gene involved with methionine metabolism and recently identified as necessary for the control of the lifespan in *Drosophila* (Parkhitko et al., 2016). Interestingly, the expression of *inx7* and *intB* did not result in any significant changes in *CapaR>GATAe* RNAi MTs (data not shown), even though these genes play a role downstream of *GATAe* in the developing MG (Okumura et al., 2005). A possible explanation of this result is provided in the discussion.

Altogether, these results indicate that *GATAe* directly or indirectly regulates the expression of several tumour-related and lifespan-related genes, and may, therefore, induce RNSC proliferation, and impact on the morphology and function of the adult MTs.

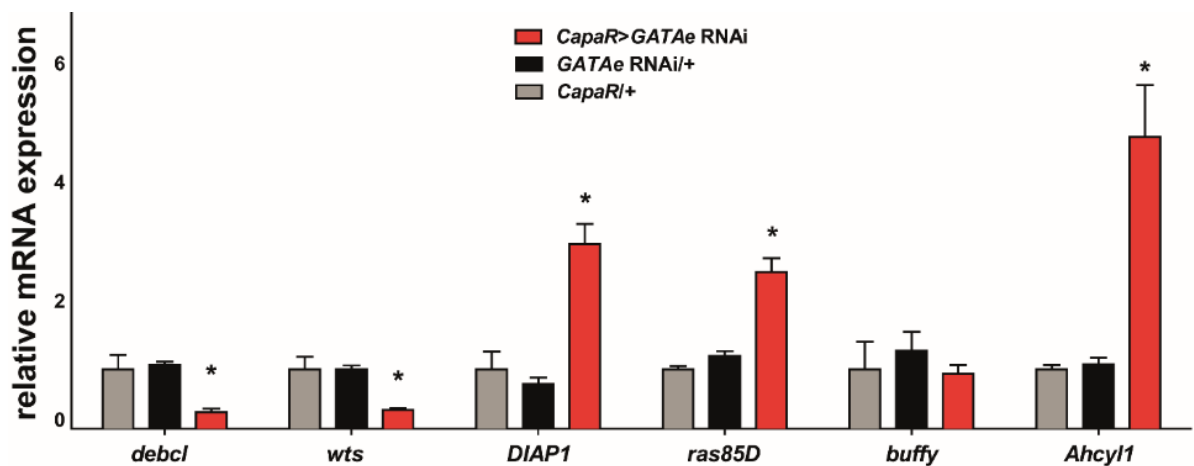


Figure 4.2.16. Relative mRNA levels of oncogenes and apoptosis-related genes in *GATAe* knockdown MTs

CapaR>GATAe RNAi MTs showed significantly increased expression of *DIAP1*, *ras85D* and *Ahcyl1*. However, *debcl* and *wts* expression was downregulated in *CapaR>GATAe* RNAi MTs compared to the controls. No changes in *buffy* expression were observed. The bars marked with * indicate significant difference ($p < 0.05$, Student t-test, two-tailed).

4.2.8 Knockdown of *GATAe* in a small number of PCs is enough to induce abnormal tissue growth

Mosaic analysis has been widely used in *Drosophila* to investigate the effects of a complete loss of a gene in a specific tissue of an otherwise wild-type animal. A widely-used technique for Mosaic analysis is that with a repressible cell marker MARCM (Lee and Luo, 1999)) and is a potent tool to investigate, e.g. the RNSCs (Singh et al., 2007, Li et al., 2015, Zeng et al., 2010). However, an important limitation of this technique is that requires mitotic recombination to trace and label the progeny, which cannot be applied to MT PCs, as this already differentiated cell population is post-mitotic (Jung et al., 2005).

As a potential alternative to study clonal genetic alteration in the PCs, the *Urate Oxidase-Gal4:UAS-CD8-GFP* (*UrO>GFP*) strain has been used in this project. The pattern of expression of the native *UrO* gene is exclusively restricted to all the PCs of the main segment (Terhzaz et al., 2010). However, in adult female flies, *UrO>GFP* is only expressed in a subset of PCs of the main segment (GFP^+ cells in Figure 4.2.17A, A'). This expression profile makes this GAL4 line an ideal way to investigate the effects of genetic manipulation in a small group of cells.

Therefore, *UrO>GFP>GATAe* RNAi adult female MTs were examined in the search for morphological defects. They exhibited critical shape abnormalities and tissue growth only in their main segments, almost exclusively in the areas surrounding the cells in which *GATAe* RNAi expression was induced (GFP^+ in Figure 4.2.17D). However, no evident increase of RNSCs has been observed in these areas, but, in some cases, GFP^+ cells, or neighbouring cells became binucleated (Figure 4.2.17E and F), suggesting endoreduplication, a classical phenotype of cancerous cells (Shi and King, 2005, Storchova and Pellman, 2004). Altogether, these experiments with *UrO-Gal4* showed that *GATAe* could act in an autonomous and non-autonomous way (affecting the morphology and behaviour of the same and neighbouring cells).

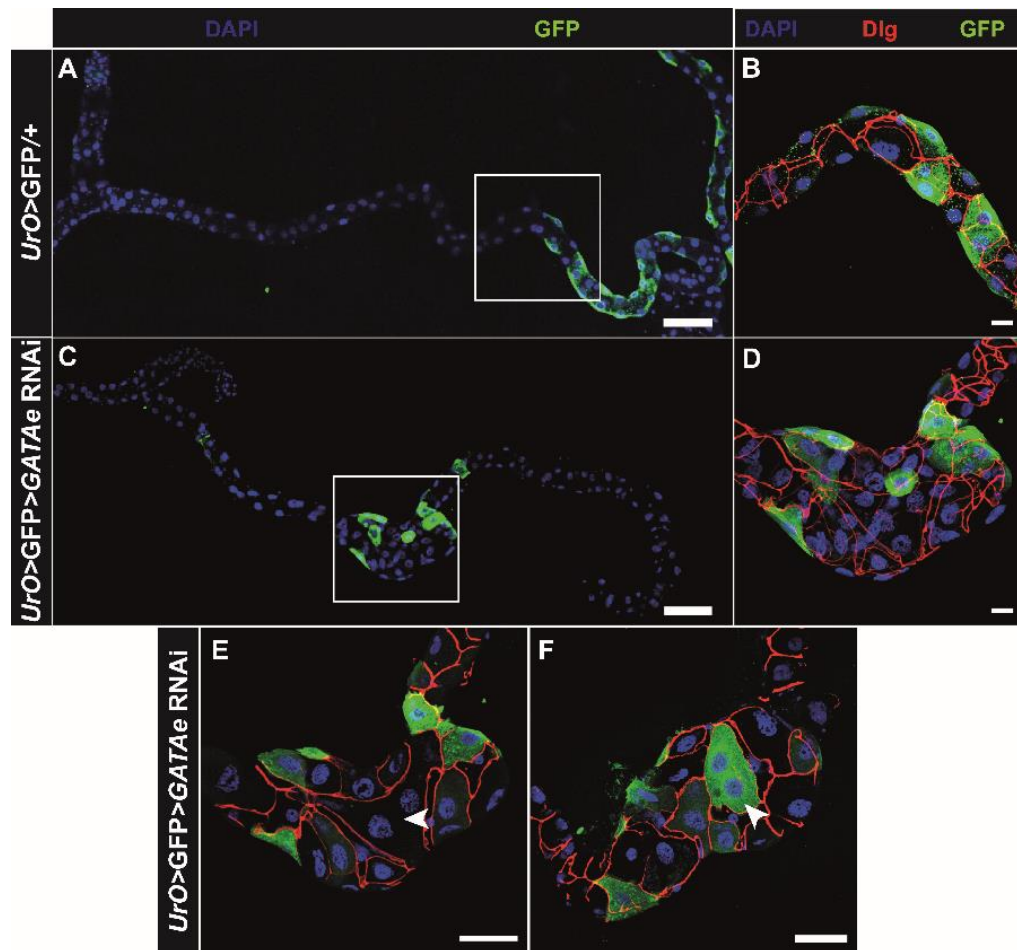


Figure 4.2.17. *GATAe* knockdown using *UrO>GFP*.

Control (A and B) and knockdown (C and D) adult female MTs. *GATAe* is downregulated exclusively in the GFP⁺ cells. B and D are magnifications of the regions boxed in white in A and C, respectively. E and F are examples of binuclear cells present in different *UrO-GFP>GATAe* RNAi main segments. Scale bars are 100µm in A and C, 20µm in B and D and 50µm in E and F.

4.3 Functional characterisation of *GATAe*

It has been shown that the MTs regulate water homeostasis, and ionic and acid-base balance, similar to the human kidney (Beyenbach et al., 2010, Dow, 2012). Impairments in MT function in any of these processes severely affects the organism survival and resistance to external stresses (Dow and Romero, 2010, Dow and Davies, 2003). Therefore, after elucidating that *GATAe* is crucial for the morphology of adult MTs, the impact of silencing *GATAe* on MT function was assessed.

In the absence of any kind of stress, 7-day-old *CapaR>GATAe* RNAi flies exhibited a bloated abdomen compared to parental control flies (Figure 4.3.1A). Interestingly, this phenotype has been previously associated with MT failure, and more specifically, an impairment in the regulation of water homeostasis (Cabrero et al., 2014, Denholm et al., 2013). Therefore, to identify if this

abdominal bloating is a consequence of water retention or either air or fat accumulation, flies were subjected to wet-dry weight measurements. Results show that wet weights of *CapaR>GATAe* RNAi flies were significantly higher compared to both parental controls. However, dry weights of both *CapaR>GATAe* RNAi and control flies were similar (Figure 4.3.1B). These results indicate that *CapaR>GATAe* RNAi flies accumulated more water compared to control flies, presumably due to impairment of tubule function.

Impaired MT function has been previously reported as a crucial factor for fly viability, stress and survival (Terhzaz et al., 2015b, Denholm et al., 2013). Furthermore, lifespan assays demonstrated that *CapaR>GATAe* RNAi flies exhibited significantly shorter lifetimes under normal conditions compared to parental control flies (Figure 4.3.1C), indicating that impairment in tubule function affects not only the MTs, but also the overall fitness of *GATAe* knockdown flies. To assess their tolerance of different kinds of stress, *CapaR>GATAe* RNAi flies were exposed to both starvation (no food) and desiccation (no food or water) stresses. Reduced tolerance to these stressors have been previously associated with impaired function of the MTs (Terhzaz et al., 2015b, Cabrero et al., 2014, Davies et al., 2014b). As expected, *CapaR>GATAe* RNAi resulted in significantly reduced survival under starvation (Figure 4.3.1E). Surprisingly, these flies were also more sensitive to desiccation stress (Figure 4.3.1D). This result could seem contradictory given that these flies also accumulate significantly more water in normal conditions (with no stress induced), but this could be a consequence of completely impaired renal functions of the MTs, inducing more sensitivity to stress in general. Finally, assessing fluid secretion rates in *CapaR>GATAe* RNAi MTs by the Ramsay secretion assay (see 2.9.4, (Dow et al., 1994)) to study their responses to neuropeptide stimulation was impossible due to their reduced size.

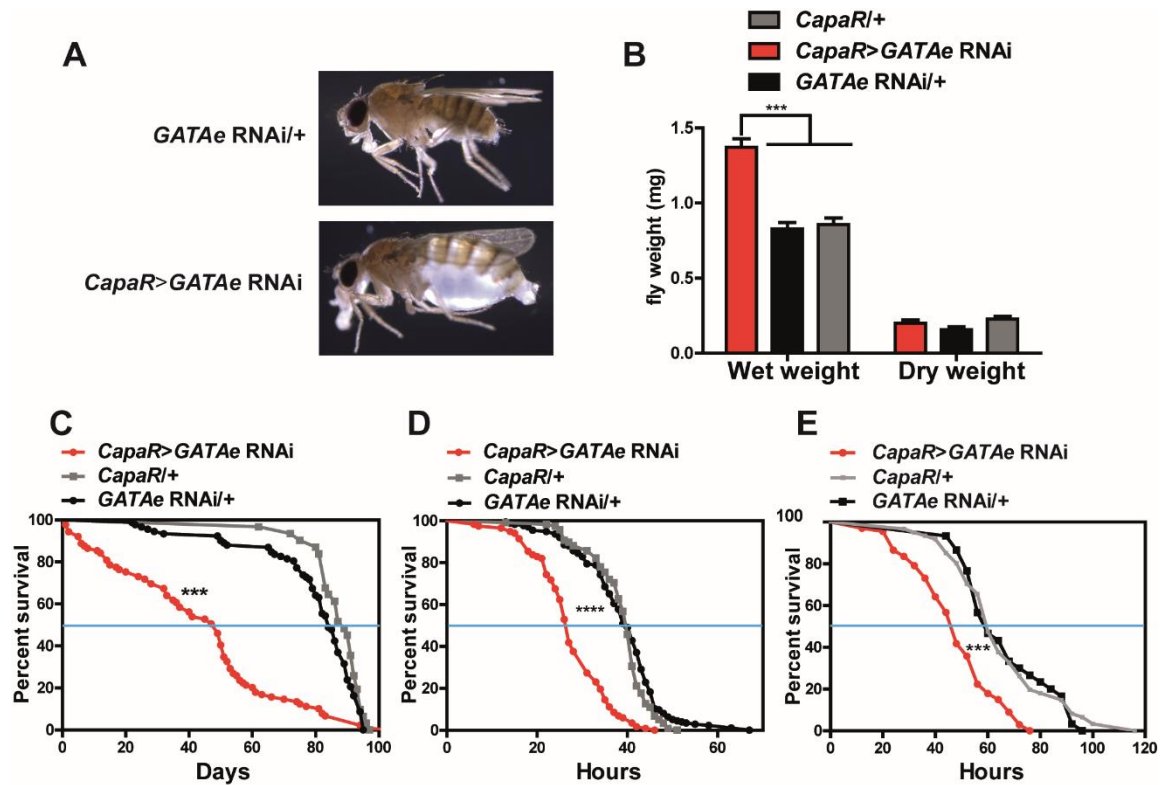


Figure 4.3.1 Reducing *GATAe* expression in PCs affects water homeostasis and stress tolerance.

(A) *CapaR>GATAe RNAi* adult flies exhibit bloated abdomens compared to the controls. (B) Wet and dry weights measurements. *GATAe* knockdown flies showed a significant difference compared to both parental controls. $P < 0.0001$, $n > 90$). (C-E) *CapaR>GATAe RNAi* flies exhibit a significant decrease of (C) lifespan, (D) desiccation stress and (E) starvation stress. Median survivals are (indicated by blue lines): (C), *CapaR>GATAe RNAi* = 49, *CapaR/+* = 89, *GATAe RNAi/+* = 84. $P < 0.0001$, $n > 100$ flies). (D), *CapaR>GATAe RNAi* = 27, *CapaR/+* = 40, *GATAe RNAi/+* = 40. $P < 0.0001$, $n > 100$ flies). (E), *CapaR>GATAe RNAi* = 48, *CapaR/+* = 60, *GATAe RNAi/+* = 60. $P < 0.0001$, $n > 100$ flies).

To summarise, these results confirm that the impaired morphology induced by silencing *GATAe* in the PCs are associated with severe deficiencies in the function of the MTs, consequently affecting the survival under environmental stress and the overall lifespan of the animal.

4.4 Discussion

Here, the cell-specific involvement of *GATAe* to maintain the proper morphology of the MTs has been studied and elucidated, and it could be confirmed that reduced levels of this gene in tubule PCs severely impacts in adult fly survival.

Surprisingly, *GATAe* is not required for embryonic development of the MTs but plays a key role during metamorphosis and in the maintenance of the adult MTs. Most of the research has focused on the development of the MTs during embryonic development (Denholm, 2013, Hatton-Ellis et al., 2007, Beyenbach et

al., 2010). The MTs remain fairly intact through all the development and even survive the steroid-regulated death and rebuilding of the MG during metamorphosis (Baehrecke, 2000). Interestingly, Wessing and Eichelberg identified significant remodelling of the MTs during the pupal stage (Wessing and Eichelberg, 1979). They found that the larval MTs, which at this stage are very similar to the adult tubules regarding tubule length and cell organisation, shrink approximately half their length during the event of metamorphosis. The MTs eventually elongate back to their original length before adult eclosion (Wessing and Eichelberg, 1979). However, little is yet known regarding which factors are necessary for their maintenance during the pupal stage. The mechanisms underlying this process of MTs changes are poorly understood, but one could think that renal functions are shut down during pupariation as no feeding takes place during this stage, allowing these critical changes in shape. Given that *GATAe* is especially required from this stage, it is likely that it plays a role in regulating this process of shrink or subsequent elongation.

In addition, the experiments performed in this part of the thesis have uncovered new roles for *GATAe* in the PCs, that consequently affect the RNSC population. It has been previously demonstrated that RNSCs originate from AMPs and migrate from the AMG to the MT ureters during metamorphosis and these migration events require the function of the small GTPase Rac1 (Takashima et al., 2013). Also, most of the studies of RNSC behaviour and maintenance have been performed predominantly in adult stages. Any disruption in gene expression in RNSCs leads to tumour formation and strong defects in the function of the MTs (Singh et al., 2007, Li et al., 2015, Bohère et al., 2018). Nevertheless, little is known about the function of the differentiated cell types of the MTs (PCs or SCs) could have to control the RNSC population. The results in this chapter show, for the first time, that *GATAe* is required in the differentiated PCs to maintain the correct architecture of the adult MTs. Strikingly, *GATAe* is also required in the PCs to maintain the right localisation and proliferation of the neighbouring RNSCs. If *GATAe* is indeed required for RNSC proliferation control, or if this RNSC proliferation is a consequence of *GATAe* knockdown in the PCs due to an indirect effect of tissular damage is a highly exciting line of research that should be done. Experimentation measuring DNA damage and modulating the expression of DNA repair proteins (e.g. p53) or regulators of cell cycle progression, such as the

Mitogen-activated kinase (MAPK) pathway, could explain the origin of these phenotypes.

One crucial point that has been difficult to address in the experiments performed in this thesis is the confirmation of an efficient knockdown of *GATAe* expression levels using a *GATAe* RNAi line. As shown in Figure 4.2.2D, the levels of downregulation obtained using the driver *GATAe*-Gal4 are relatively low (only 40% knockdown approximately). However, it has been difficult to obtain an evident knockdown for *GATAe* using any other MT-specific Gal4 line.

There may be several reasons for this. First, all the Gal4 strains used to silence *GATAe* most probably still produce residual expression in neighbouring tissues. Several Gal4 lines have been used to confirm *GATAe* knockdown: *CapaR*-Gal4 is expressed in the PCs, but not in SCs or RNSCs. Therefore, and as mentioned in the following chapters, expression of *GATAe* in SCs or RNSCs could potentially mask low levels of *GATAe* in the PCs (especially in the adult stage, as silencing *GATAe* in the PCs induces overproliferation of RNSCs, as shown in 4.2.6). qPCR experiments have also been performed in MGs with reduced levels of *GATAe* (using *esg*-Gal4), and very slight variation in *GATAe* expression was observed. In this case, *GATAe* expression in neighbouring cells (Zhai et al., 2018, Dutta et al., 2015, Okumura et al., 2016, Dobson et al., 2018) could also mask any induced expression in target cells. Also, the driver *Actin*-Gal4 has been employed as an attempt to downregulate *GATAe* ubiquitously. As mentioned before, this resulted in premature lethality shortly after embryogenesis. Therefore, RNA was extracted from whole L1/L2 specimens for subsequent qPCR. However, this approach did not result in a conclusive reduction of *GATAe* expression levels either, suggesting that even when using *Actin*-Gal4 as a driver, some residual expression of *GATAe* can remain. This could be solved by using a thermo-sensitive *Actin*-Gal4 to knockdown *GATAe* only from the adult stage, and perform RT-PCR in isolated MTs.

Nevertheless, the phenotypes and results presented upon *GATAe* knockdown in this thesis are entirely reliable, for several reasons. First, four independent RNAi lines used here (BDSC #33748 and #34907, and VDRC #10420 and #10418) induced the same morphological abnormalities when driven in the PCs (Figure 4.2.5), with slight differences in survival (e.g. driving the *GATAe* RNAi line #34907 using

either *CapaR* or *ctB*-Gal4 resulted in premature lethality before adult eclosion). Second, all these RNAi lines have been previously employed to knockdown *GATAe* in other tissues in multiple publications by independent research groups (e.g. Okamura et al., 2016 utilised BL33748, V10420 and V10418, Dutta et al., 2015 made use of BL33748, and both Zhai et al., 2017 and Zhai et al., 2018 utilised BL34907). Third, all the PC-specific Gal4 lines employed in this chapter (see section 2.1.2) also produced similar phenotypes when combined with *GATAe* RNAi. These PC-specific Gal4 lines include *CapaR*-Gal4 (Terhzaz et al., 2012), *c42*-Gal4 (Broderick et al., 2004), *ctB*-Gal4 (Saxena et al., 2014) and *UrO*-Gal4 (Terhzaz et al., 2010). Also, the phenotypes obtained using the *GATAe*-Gal4 lines (Figure 4.2.4) were similar to the ones using specific PC lines, and phenocopied the MG abnormalities reported in previous reports (Okumura et al., 2005, Okumura et al., 2016).

Therefore, all the combination of RNAi and Gal4 lines employed in this thesis confirm not only that the phenotypes observed are a consequence of specific silencing of *GATAe*, but also that this downregulation is limited to the MT PCs.

Altogether, and as commented above, the already differentiated PCs require the expression of *GATAe* to ensure the tubule identity and function in metamorphosis and adult stage, and to control RNSC proliferation. Low levels of *GATAe* in the PCs, therefore induce severe defects in morphology, which in consequence impact MT function. Thus, these data open a new window in the investigation of the development of the MTs and could give rise to the identification and study of other factors necessary in this organ during pupal or adult stages.

Two main hypotheses could explain the relationship between PC and RNSC populations. In the first scenario, *GATAe* would be essential for the maintenance of the PC population. Therefore, PCs with low levels for *GATAe* would not survive and start showing hallmarks of cancer, which losses the control of RNSC proliferation as they try to replenish the PC population. In other words, these RNSCs would migrate and proliferate to respond to tissue damage caused by *GATAe* knockdown in the PCs. This first hypothesis presents important limitations. In most *CapaR*>*GATAe* RNAi MTs observed, there is the presence of RNSCs in regions where they usually are not present (e.g. in the main segment),

but they very rarely reach the more distal regions such as the initial segment of the tubule. A possible explanation of why RNSCs gather and don't seem to differentiate could be that in fact, they go through a cycle of differentiation-death. *GATAe* RNAi PCs would go through apoptosis, inducing RNSCs to migrate and differentiate into new PCs. However, these already differentiated PCs would also start expressing the *GATAe* RNAi construct and start dying, repeating this cycle of RNSC differentiation and posterior PC death.

Also, this explanation does not provide a clear answer for why RNSCs overproliferated under *GATAe* knockdown conditions, and, more importantly, why these RNSCs are not able to differentiate to PCs and end up accumulating in almost all regions of the MTs. In addition, this hypothesis does not give an explanation as for why RNSCs are not proliferating upon *GATAe* downregulation using *UrO*-Gal4 in a similar way as they do when using *CapaR*-Gal4.

A more plausible explanation is that expression of *GATAe* in the PCs plays a vital role in the control of the RNSC population (see next chapter). Expression of *GATAe* in the PCs would be required for signalling to adjacent cells required to block subsequent migration and proliferation of RNSCs, thus regulating critical apoptotic and cancer-related genes, directly or indirectly. This second hypothesis would also explain why there are no RNSCs present in the areas affected by *GATAe* knockdown using *UrO*-Gal4 (Figure 4.2.17), as the pattern of expression of *UrO*-Gal4 does not colocalise with the region where RNSCs reside. In consequence, and according to this second theory, the RNSC population is not affected by the lack of *GATAe* exclusively in the main segment.

However, even if this second hypothesis explains the phenotype observed upon *GATAe* downregulation in the PCs, further experimentation needs to be performed to precisely identify the nature of the precise role of *GATAe* in the PCs. A proposed model of *GATAe* roles in the MTs is presented in 5.5.

4.4.1 Evolutionarily conserved functions of the GATA family

The findings of this chapter represent a strong link in GATA factor functions along evolution and are in line with the immense amount of research that has been performed to understand the role of GATA factors in humans, especially in

the field of cancer. Indeed, mutations or alterations in the expression of almost all GATA factors in humans have been associated with diverse types of cancers. The specific roles of GATA factors in cancer can vary strongly depending in their cellular context, as the same GATA factor can act both as tumour suppressors or oncogenes in different types of cancer (see 1.4.2.2). Therefore, and in accordance to the results shown in this thesis (in this chapter and in Chapter 5), diverse functions of GATA factors depending on their cellular context seem to be evolutionary conserved.

The gene expression data reported here (especially the altered levels of *DIAP1* and *debcl* in PC-specific *GATAe* downregulation) further supports the hypothesis of *GATAe* as a master regulator of proliferation and growth in the MTs.

Downregulation of *wts* in *CapaR>GATAe* RNAi MTs is in line with previous data that showed that RNSCs mutant for *wts* lose their proliferation control and do not stop dividing (Zeng et al., 2010). Thus, it is very likely that *GATAe* knockdown in the PCs results in loss of *wts* in the RNSCs. The altered levels of *debcl* also correlate with the phenotype observed of abnormal proliferation. Several research groups have reported interactions between GATA factors and members of the bcl-2 family in vertebrates. In mice, GATA1 interacts with bcl-x to ensure the survival erythroid cells (Gregory et al., 1999), and GATA4 can also bind directly to another member of the bcl family, Bcl2 (Kobayashi et al., 2006, Aries et al., 2004, Kobayashi et al., 2010), to induce cell survival due to drug-induced toxicity in the heart. Therefore, it is very likely that in *Drosophila*, an interaction between GATA factors (*GATAe*) and bcl-2 members (*debcl*) also takes place to ensure homeostasis of the MTs. Although it has not been precisely investigated in this thesis, it is also likely that the differences in expression of *DIAP1* and *debcl* (as well as *wts*, mentioned above), come from the RNSCs. This has been supported by previous data, that *DIAP1* is expressed in RNSCs (Bohère et al., 2018), and that *debcl* knockdown does not seem to alter survival in PCs in response to oxidative damage (Terhzaz et al., 2010).

It would be compelling to perform further research to identify the exact cell-type in the MTs responsible of these specific changes of gene expression. This would help to identify the pathways that modulate the striking morphological defects found in this thesis.

To summarise, the overall findings indicate that *GATAe* controls homeostasis of the MTs, and silencing this gene induces critical deficiencies, inducing tumorigenic growth and loss of proliferation control. Therefore, it could potentially act as a tumour suppressor, in a similar way as its counterparts in vertebrates. Thus, these results open a new window to establish and study homologies between insect and human GATA factors and provide evidence that the mechanisms of tissue homeostasis can be potentially conserved along evolution.

Chapter 5 - Analysis of *GATA*e in the Stellate and Renal and Nephritic Stem Cells

5.1 Introduction

The work described in previous chapters has demonstrated that *GATAe* plays a vital role in the PCs of pupal and adult MTs and is required for functional physiology of MT and also organismal survival. This chapter focuses on the function of *GATAe* in the other two cell populations of the MTs: SCs and RNSCs. The results obtained by silencing *GATAe* specifically in these cell populations showed that although the morphological defects are not as severe as silencing *GATAe* in PCs, *GATAe* is still required for SC survival, which impacted on the neuropeptide-induced MT fluid secretion response. Early migration of RNSCs into the ureter is also affected by *GATAe* silencing, but not their survival during the adult stage.

5.2 *GATAe* in Stellate Cells

The driver *CIC-a*-Gal4 (Cabrero et al., 2014) was used in this study to silence *GATAe* in the SCs specifically. Unlike *CapaR*>*GATAe* RNAi MTs, *CIC-a*>*GATAe* RNAi MTs did not show any defect in the length or the overall structure of the MTs. However, *CIC-a*>*GATAe* RNAi adult MTs exhibited a significant reduction in SC number (Figure 5.2.1C). Moreover, the reduced number of detected SCs were mostly confined to the initial segment of the MTs (especially in the anterior MTs, in which bar-shaped cells are located), and almost absent in the main segment where SCs normally reside (Figure 5.2.1B).

Next, Ramsay secretion assays were performed on *CIC-a*>*GATAe* RNAi MTs to investigate physiological function. The *Drosophila* leucokinin (herein referred as kinin) neuropeptide specifically stimulates intracellular $[Ca^{2+}]$ in the SCs, leading to the activation of Cl^- conductance into tubule lumen and therefore, fluid secretion (Terhzaz et al., 1999, O'Donnell et al., 1998, Cabrero et al., 2014). Therefore, *Drosophila* kinin was used to test if the reduced number of SCs affect basal or stimulated fluid secretion in *CIC-a*>*GATAe* RNAi MTs.

As expected, fluid secretion assays demonstrated that while basal secretion rates did not change significantly compared to control MTs, kinin-stimulated secretion rates in *CIC-a*>*GATAe* RNAi MTs were significantly lower compared to

both controls (Figure 5.2.1D). These results indicate that *GATAe* is required for the maintenance of the SC population in the main segment.

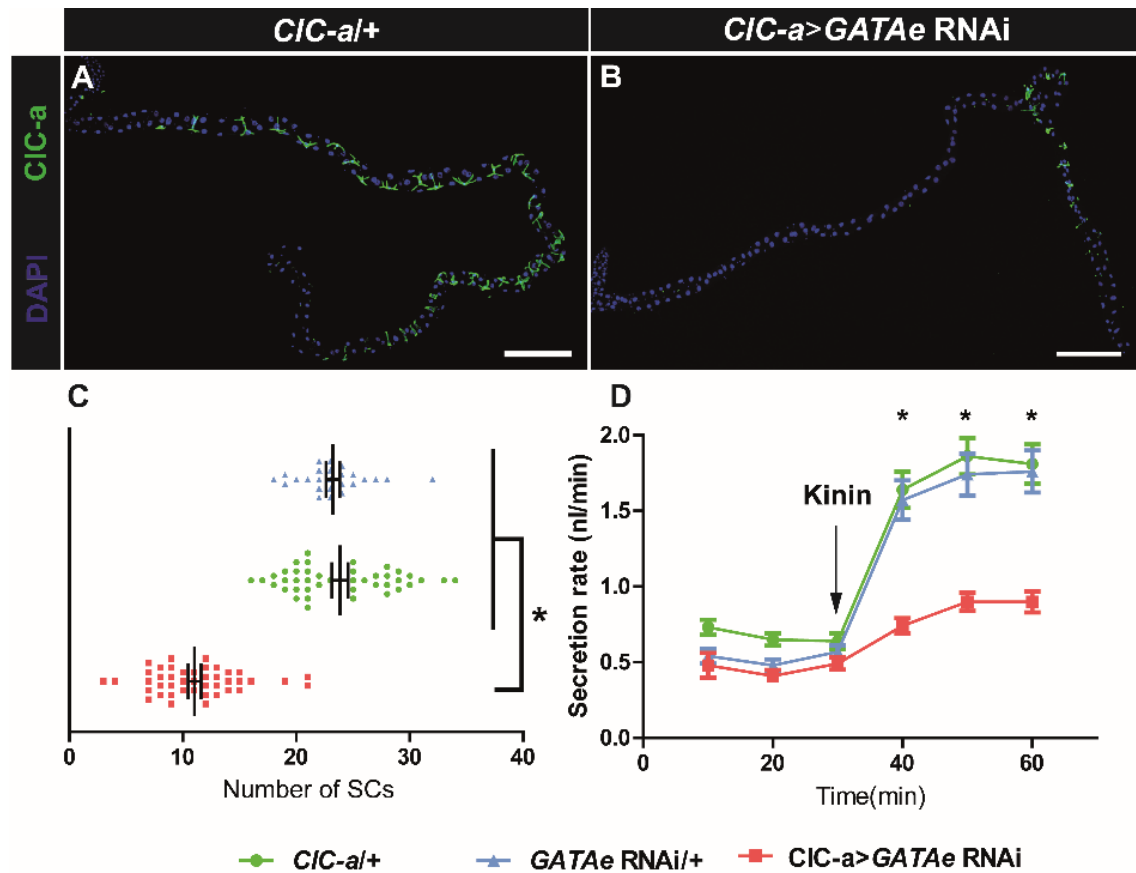


Figure 5.2.1. *GATAe* downregulation in the SCs induces loss of the SC population and reduced fluid secretion response to kinin stimulation in the adult stage.

(A, B) ICC of adult control (A) and *GATAe* knockdown in the SCs using *CIC-a-Gal4* (B). *CIC-a>GATAe RNAi* MTs show reduced number of SCs, which are confined to the initial segments. Scale bars are 200 μ m. (C) SC number in adult female tubules. *CIC-a>GATAe RNAi* MTs (red) show reduced number of SCs compared to the controls (green and blue). Data are presented as mean \pm SEM * $P < 0.05$. Means are: *CIC-a/+* = 23.86 \pm 4.6, $n=42$, *GATAe RNAi/+* = 23.24 \pm 3, $n=25$, *CIC-a>GATAe RNAi* = 11.04 \pm 3.8, $n=45$. (D) Fluid secretion rates (nL/min⁻¹) in control and *GATAe* knockdown Malpighian tubules. *Drosophila* kinin was added at 30 min (arrow). Response to 10⁻⁷ M kinin is significantly reduced in *GATAe* knockdown tubules (red), compared with both control conditions (green and blue). Significance is measured with Student-t-test, two-tailed, $P < 0.05$.

During embryogenesis, SCs go through a mesenchymal-to-epithelial transition, resulting in their intercalation into the tubule. These transition and integration processes require expression of polarity proteins in the embryonic PCs to allow proper SC intercalation (Campbell et al., 2010). To test the possibility that *GATAe* could be involved in any of these events, *CIC-a>GATAe RNAi* MTs were observed in the larva. At wandering L3 stage, *CIC-a>GATAe RNAi* MTs showed a considerably higher number of SCs compared to the adults (16.42 \pm 2.5 SCs in L3 stage vs 11.04 \pm 3.8 SCs in adult stage), but still significantly lower than L3 controls (Figure 5.2.2E). However, SCs in wandering L3 MTs were distributed

along the main and initial segments in the standard pattern (Cabrero et al., 2014)(Figure 5.2.2C), suggesting that SCs adequately integrate into the tubule during embryogenesis. However, *clc-a* is only expressed in embryonic SCs after they integrate into the tubule (Denholm et al., 2013). Thus, even though *GATAe* is clearly necessary for SC survival during metamorphosis, a role for *GATAe* in SC integration in the MTs during embryogenesis cannot be excluded.

Remarkably, the *Clc-a* antibody used here to mark SCs did not stain the SCs of the initial regions of adult anterior MTs, corresponding to “bar-shaped” cells (Sözen et al., 1997), in both control and knockdown conditions (Figure 5.2.2B and D). This suggests that prospective L3 bar cells don’t express *Clc-a* until the adult stage. Therefore, it would indicate that *Clc-a>Gal4* would not drive expression in all SCs (at least in L3 stage), and, potentially, that the presence of *Clc-a*⁺ cells in *Clc-a>GATAe* MTs could be due to the lack of expression of this driver in SCs of the initial segment.

Altogether, the data presented here indicate that *GATAe* plays a key role in SC maintenance and function, but it is not required for SC integration or intercalation into the embryonic MTs.

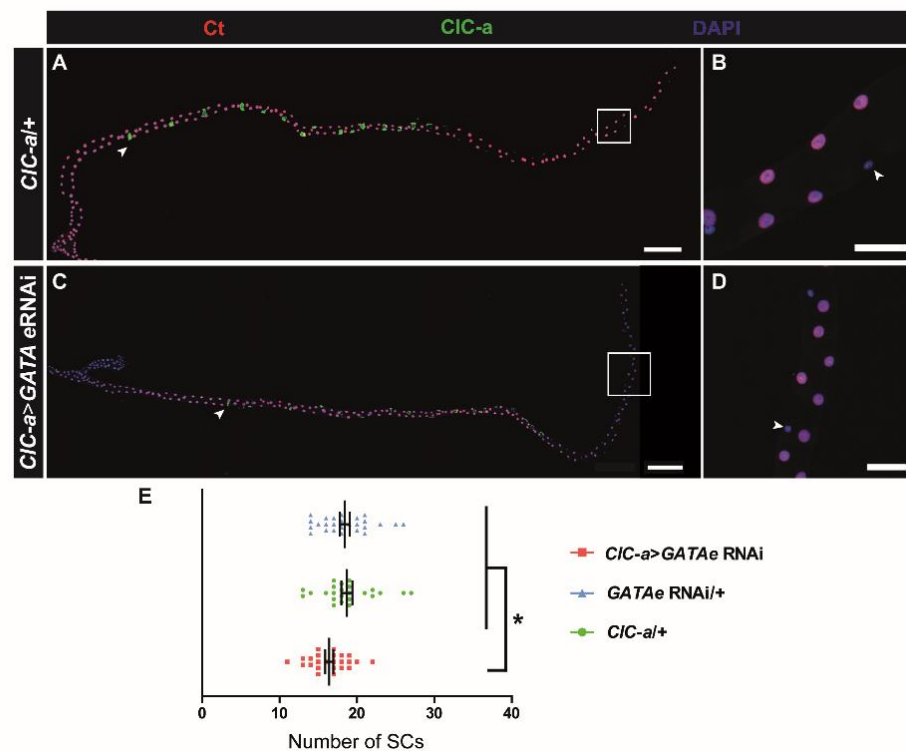


Figure 5.2.2. SCs show normal localisation in the L3 stage in *GATAe* knockdown conditions.

(A, D) ICC of L3 control (A and B) and *CIC-a>GATAe* RNAi MTs (C and D). *CIC-a>GATAe* RNAi MTs show normal localisation of SCs (white arrows in A and C). B and D are magnifications of the white squares in A and C, respectively, showing SCs negative for *CIC-a* and *ct* (white arrows), which are prospective bar cells. Scale bars are 200µm (A and C) and 50µm (B and D). (E) SC quantifications in adult female MTs. *GATAe* knockdown MTs (red squares) exhibit a small but significant reduction in their SC number compared to the controls (green circles and blue triangles). Data are presented as mean \pm SEM * $P < 0.05$. Means are: *CIC-a/+* = 18.71 ± 3.48 , $n=24$, *GATAe* RNAi/+ = 18.42 ± 3.28 , $n=24$, *CIC-a>GATAe* RNAi = 16.42 ± 2.5 , $n=26$.

5.3 *GATAe* in the RNSCs

In addition to the involvement of *GATAe* in PCs and SCs maintenance, a possible role of *GATAe* in the migration, survival, and maintenance of RNSCs was investigated. For this, the *esg*-GAL4-UAS-GFP driver combined with a Gal80^{ts} construct (*esg>GFP^{ts}*), which induces specific temporal GAL4 expression to both the ISCs and RNSCs, (Singh et al., 2007, Micchelli and Perrimon, 2006) was used in this project. As previously shown using the temperature inducible expression Gal80^{ts} system (see 4.2.5), flies were kept at 18°C until different developmental stages (wandering L3, 24h APF or adult stage) and then switched to 29°C to induce GAL4 expression, thereby silencing *GATAe* expression. The RNSC population of these flies was analysed in the adult stage.

Interestingly, silencing *GATAe* expression specifically in the RNSCs in wandering L3 stage resulted in an almost complete absence of RNSCs in the adult ureters (Figure 5.3.1A'). In contrast, when *GATAe* expression was downregulated from 24h APF, the stage at which the RNSCs have already initiated their migration to the MTs (Takashima et al., 2013) this resulted in a normal pattern, but reduced number of RNSCs in the ureters (Figure 5.3.1B').

In addition, as *GATAe* is crucial for MG ISC/EB maintenance (Okumura et al., 2016), could it also be necessary for adult RNSC maintenance? To answer this question, *GATAe* expression was silenced in RNSCs in newly emerged flies. At this stage, RNSCs have completely migrated to the ureters and lower tubules (Singh et al., 2007). After 20 days of *GATAe* downregulation using *esg>GFP^{ts}>GATAe* RNAi, there was no difference in the presence of RNSCs in ureters and lower ureters (Figure 5.3.1C and C').

Therefore, these findings show that *GATAe* is potentially required for the early migration of the RNSCs, but when they have started to migrate, *GATAe* either

seems not to be required, or other factors can complete the RNSCs migration. Also, they show that adult RNSCs do not require *GATAe* for their survival.

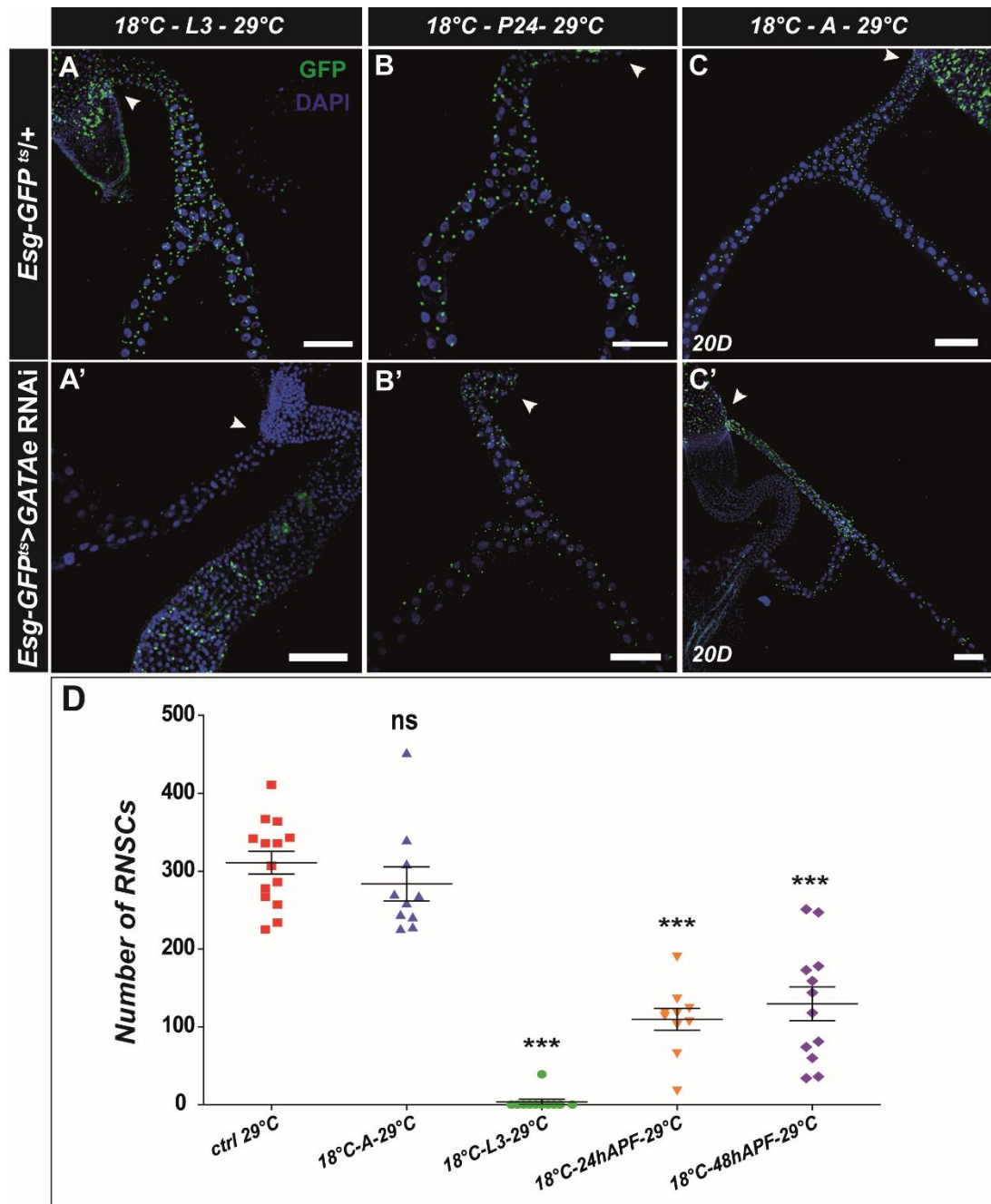


Figure 5.3.1. RNSC migration to the MTs is impaired upon *GATAe* knockdown.

GFP staining of adult ureters using *esg>GFP^{ts}* driver. (A, A') Flies were raised at 18°C until wandering L3 stage and then switched to 29°C. *Esg>GFP^{ts}>GATAe RNAi* tubules show no presence of RNSCs compared to the controls. only a few exceptions displayed the presence of RNSCs, but no more than 40 in any case (higher dots in D). (B, B') Flies were raised at 18°C until 24h APF and then switched to 29°C. *Esg>GFP^{ts}>GATAe RNAi* tubules under these conditions have normal RNSC population in their ureters. (C, C') Flies that were raised at 18°C until adult eclosion and then switched to 29°C for 20 days present normal localisation of RNSC population in *esg>GFP^{ts}>GATAe RNAi*, compared to the controls. Scale bars are 100µm. Arrowheads indicate the ureters. (D) *Esg⁺* quantifications of 7-day-old adult ureters (using ImageJ), which have been transferred from 18°C to 29°C at the L3 stage. Means are: *Esg-GFP^{ts/+}* = 297.9 cells ± 20.03 SEM, n=8, *Esg-GFP^{ts/+}>GATAe RNAi* = 6.250 cells ± 4.218 SEM n=12. *** indicate statistical significance (P<0.0001, Student-t-test, two-tailed).

Furthermore, these findings were confirmed using an alternative *esg>GFP* line (without the Gal80^{ts} construct). The localisation of RNSCs was examined in *esg>GFP>GATAe* RNAi MTs during different developmental stages. In the wandering L3 stage, *esg*⁺ cells were found in clusters in the MG in both control and knockdown conditions (Figure 5.3.2D and H). However, no RNSCs were found in *esg>GFP>GATAe* RNAi ureters at 48h APF (Figure 5.3.2P), which further reinforces the results observed using *esg>GFP^{ts}* line.

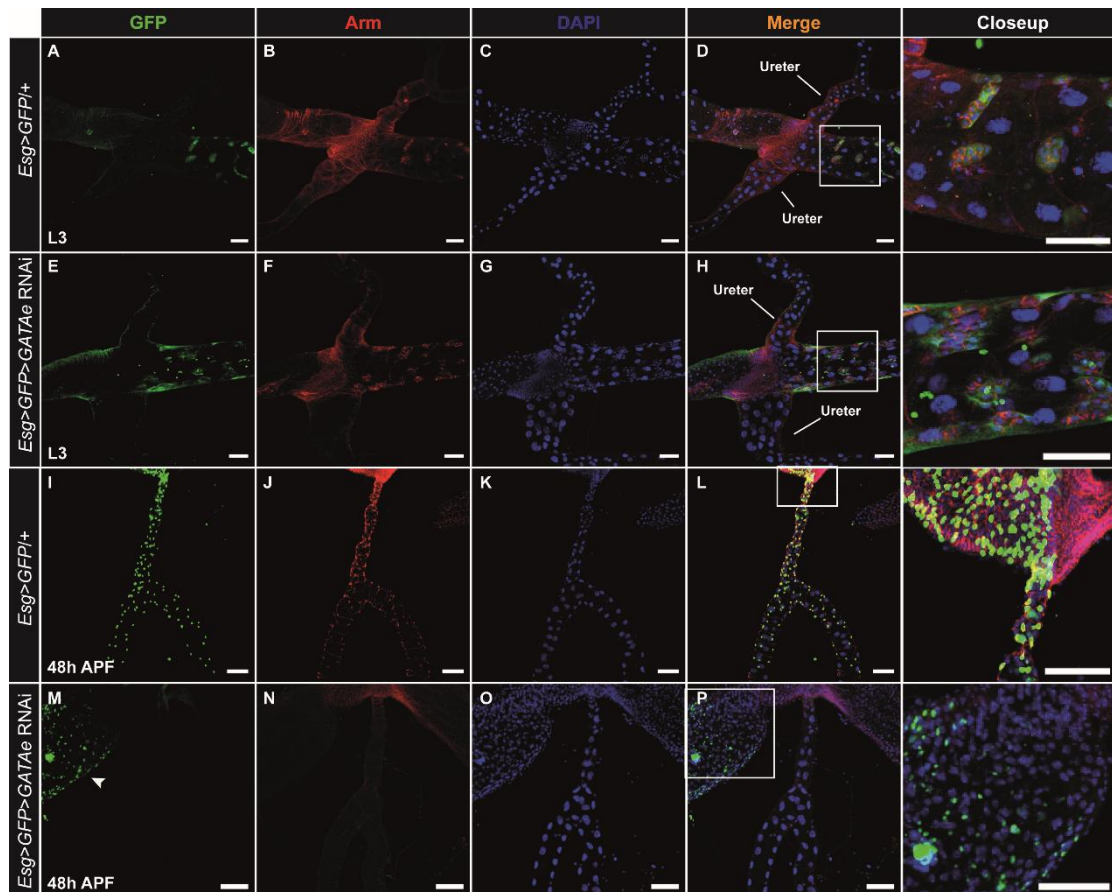


Figure 5.3.2. Migration defects of RNSCs in *GATAe* knockdown MTs.

ICC of L3 (A-H) and 48h APF (I-P), marked with GFP (green, *esg*⁺ cells), Arm (red, RNSC membranes) and DAPI. *Esg>GFP>GATAe* RNAi MGs show normal localisation of *esg*⁺ cells, which are clustered (arrows in D vs H). However, RNSCs do not migrate to the ureter as in 48h APF *esg>GFP>GATAe* RNAi MTs (P), whereas RNSCs in control MTs have already been relocated to the ureters (L). Closeup images represent the areas highlighted with a white square in the merged MGs. Scale bars are 50µm.

5.4 Discussion

In this chapter, the function of *GATAe* in SCs and RNSCs was investigated. The results obtained here show that *GATAe* function is not only crucial in the PCs, but it is also necessary for the survival of SCs and possibly the early migration of the RNSCs.

5.4.1 *GATAe* and SC population

The data showed in this chapter clearly demonstrate that *GATAe* is required for SC maintenance. Intriguingly, the very limited number of SCs that survive in *GATAe* knockdown MTs seem to be retained to the initial segment. This could be a consequence of the expression pattern driven by *CLC-a*-Gal4, which does not seem to drive in the bar-shaped SCs in the L3 initial segment, as demonstrated by ICC using anti-CLC-a antibody (Figure 5.2.2). This is particularly evident in anterior MTs, which have a higher number of SCs upon *GATAe* knockdown compared to posterior MTs, which do not possess an enlarged initial segment. Despite this, *GATAe* knockdown MTs gradually lose most of their SCs during metamorphosis. Yet, the localisation of SCs in the larval tubule is not altered upon *GATAe* knockdown, suggesting that this cell loss might be independent from SC mesenchymal-to-epithelial transition that occurs in embryonic MTs. Previous research suggested that another GATA factor, *srp*, is expressed in SCs before their integration to the embryonic MTs, but *srp* function in this event has not been clearly elucidated (Artero et al., 2006), although a later study identified *srp* as a master regulator of epithelial-to-mesenchymal transition in other tissues (Campbell et al., 2011). It would be compelling to investigate if *GATAe* plays a role during SC integration during embryogenesis. The utilisation of GAL4 lines that are expressed in SCs during this process (e.g. *c724*-Gal4 (Denholm et al., 2003)), and the *GATAe* null mutant line would answer this question.

Remarkably, the shape of the limited number of SCs observed in *GATAe* knockdown conditions is not altered when compared to wild-type tubules. This observation is in contrast to when the TF *tsh* is silenced specifically to the SCs. It has been shown that *tsh* is necessary in SCs during pupariation to control their transition from a cuboidal to the classic stellate shape in the adult (Denholm et al., 2013). These results indicate that *GATAe* might be acting independently to *tsh*, regulating SC survival rather than their morphology. It would be interesting to investigate, in future studies, the mechanisms by which *GATAe* operates in order to maintain the SC population.

Remarkably, even though the SC population is almost abolished in *GATAe* knockdown MTs, flies do not show an inflated abdomen, as compared to *GATAe*

knockdown in PCs, or in cases where the expression of SC-specific genes (e.g. *CIC-a* or *tsh*) is disrupted (Denholm et al., 2013, Cabrero et al., 2014). Thus, although *CIC-a>GATAe* RNAi MTs display impaired kinin-stimulated diuresis, they are still capable of regulating water homeostasis while having a limited number of SCs. This leads to two interesting conclusions. First, that *GATAe* is not essential for the function of the SCs, only for their survival. Second, that the defects induced by *GATAe* knockdown in the SCs do not affect basal fluid secretion of SCs, suggesting that the few SCs left are able to functionally compensate the loss of the majority of the SC population.

5.4.2 *GATAe* and RNSC population

Results showed here demonstrate that *GATAe* is essential for the migration of RNSCs from the posterior MG to the ureter during metamorphosis. Curiously, these data also suggest that *GATAe* is not necessary for later maintenance of adult RNSC population. However, it remains to be elucidated if *GATAe* could play a role in the self-division of RNSCs. Unlike ISCs, which present higher rates of division (Micchelli and Perrimon, 2006), RNSCs are relatively quiescent (Zeng et al., 2010, Li et al., 2015), and a possible role of *GATAe* in regulating their proliferation rates could be an interesting avenue for future investigation.

The migration of RNSCs into the MTs has been thoroughly studied (Takashima et al., 2013) and discussed in the introduction (see 1.2.3.9). The small GTPase *Rac1* is required for correct RNSC migration, and early disruption of *Rac1* protein completely abolishes this process. However, inactivation of *Rac1* protein activity after 48h APF did not result in a migration defect, suggesting that *Rac1* is required only for the early stages of migration (Takashima et al., 2013). However, no other factors that control this migration between organ boundaries, to our knowledge, have been reported. Curiously, *GATAe* knockdown in RNSCs resulted in strikingly similar MT phenotypes. Therefore, it could be possible that *GATAe* and *Rac1* work in a common pathway (i.e. by interacting directly or indirectly) to orchestrate the migration of RNSCs to the pupal ureters. In addition, the fact that *GATAe* is expressed and necessary in the RNSCs could indirectly explain why the levels of two downstream genes, *inx7* and *intB*, are not altered when *GATAe* is silenced in the PCs (as mentioned in 4.2.7). If *GATAe* knockdown in the PCs induces a higher number of RNSCs (as shown in 4.2.6), it

would be possible that these cells also express *GATAe*. In consequence, *GATAe* in these proliferating RNSCs would induce the expression of both *inx7* and *intB*, therefore, masking the low levels of these genes in the PCs, and making any PC knockdown undetectable by qPCR.

Further investigation is needed to identify the precise role of *GATAe* in RNSC migration, and potential factors that mediate the early and late migration of this cell population into the MTs.

5.5 Proposed model for *GATAe* function in the MTs

To summarize, *GATAe* is necessary for all the three main types of cells of the MTs, with different functions for each cell type. It may not be surprising that *GATAe* can act differently in distinct cell populations within the same tissue. For instance, *GATAe* is expressed in all regions and cell types of the adult MG (Buchon et al., 2013, Dutta et al., 2015). While it is required in the ISCs for their self-renewal and to maintain their stemness (Okumura et al., 2016), it also determines the differentiation process in the enteroblasts, acting downstream of *Sox21a* (Zhai et al., 2017). Furthermore, *GATAe* has additional roles in the enterocytes by maintaining their morphological identity (Buchon et al., 2013), blocking their shedding following bacterial infection (Zhai et al., 2018) and modulating the effect of dietary restriction in lifespan (Dobson et al., 2018). Therefore, the results shown here are in line with previous studies and illustrate evidence of the plasticity of a TF within its specific cellular context. Also, human GATA factors can play opposite roles in different conditions as well. For instance, GATA3 can act both as a tumour suppressor and as an oncogene (see 4.4.1, and (Kouros-Mehr et al., 2008, van Hamburg et al., 2008, Nawijn et al., 2001)).

A proposed model that summarises the function of *GATAe* in PCs, SCs and RNSCs of the MTs is shown below (Figure 5.5.1). This model includes the hypothesis that *GATAe* is involved in the signalling from the PCs to the RNSCs. It suggests that the expression of *DIAP1* and *debcl* are confined to the RNSCs, and it is indirectly controlled by the expression of *GATAe* in the PCs (dotted line in Figure 5.5.1).

The qPCR data shown in this study support this hypothesis, as overexpression of *GATAe* in the PCs does not have opposite effects in gene expression compared to *GATAe* silencing (see 6.4.2). For instance, in *CapaR>GATAe*-HA MTs there is no change in the expression of *wts*, *DIAP1* or *Ras85D* (as the strong upregulation of *DIAP1* or inhibition of *wts* and *debcl* in *CapaR>GATAe* RNAi could suggest). In addition, RNSC presence or migration does not seem to be affected when *GATAe* is overexpressed in the PCs. A possible explanation is that the cells that express *DIAP1* and *debcl* in *GATAe* knockdown conditions are not the PCs, but the RNSCs (expression of *DIAP1* and *debcl*, at the same time, could be the cause of these cells not dying).

Previous publications have demonstrated that *DIAP1* is expressed in RNSCs, where it plays a vital role to ensure their survival (Bohère et al., 2018). In addition, *debcl* has been previously described as a crucial player for the induction of apoptosis (Colussi et al., 2000), and, interestingly, this gene is also strongly enriched in the MTs, according to FlyAtlas 2 and (Terhzaz et al., 2010). However, it does not seem to be involved in cell death upon oxidative stress in the PCs (Terhzaz et al., 2010). Therefore, as *debcl* is significantly downregulated when there is an excessive proliferation of RNSCs (e.g. in *CapaR>GATAe* RNAi MTs), it is likely that this gene is present in the RNSCs in normal conditions, and at endogenous levels, it ensures that RNSCs do not overproliferate. The proposed model presented in this thesis also indicates that Ras signalling pathway operates in the PCs, but also it is possible that it operates in RNSCs. The phenotypes observed by constitutively activating Ras pathway in the PCs (with *Ras^{V12}*, Figure 4.2.8) are similar to when *GATAe* is downregulated in the same cell type (Figure 4.2.7). This suggests that both pathways are acting cell autonomously. However, it has been also showed in this thesis that blocking Ras signalling in *GATAe* knockdown conditions does not rescue the growth phenotypes (Figure 4.2.8), suggesting either that Ras signalling is being altered in RNSCs or that other factors are acting downstream of *GATAe*. Finally, *Ahcyl1* is also proposed to act downstream of *GATAe*. *Ahcyl1* has been previously associated with lifespan control in already differentiated tissues, such as the fat body or the intestine (Parkhitko et al., 2016). However, the role of this gene in stem cell populations has not been yet studied, and it would be interesting to

investigate if *AhcyL1* has any implication in the other cell-types (SCs or RNSCs) of the MTs.

To summarise, the overall findings indicate that *GATAe* controls homeostasis of the MTs, and silencing this gene induces critical deficiencies, inducing tumorigenic growth and loss of proliferation control. Therefore, it could potentially act as a tumour suppressor, in a similar way as its counterparts in vertebrates. Thus, these results open a new window to establish and study homologies between insect and human GATA factors and provide evidence that the mechanisms of tissue homeostasis can be potentially conserved along evolution.

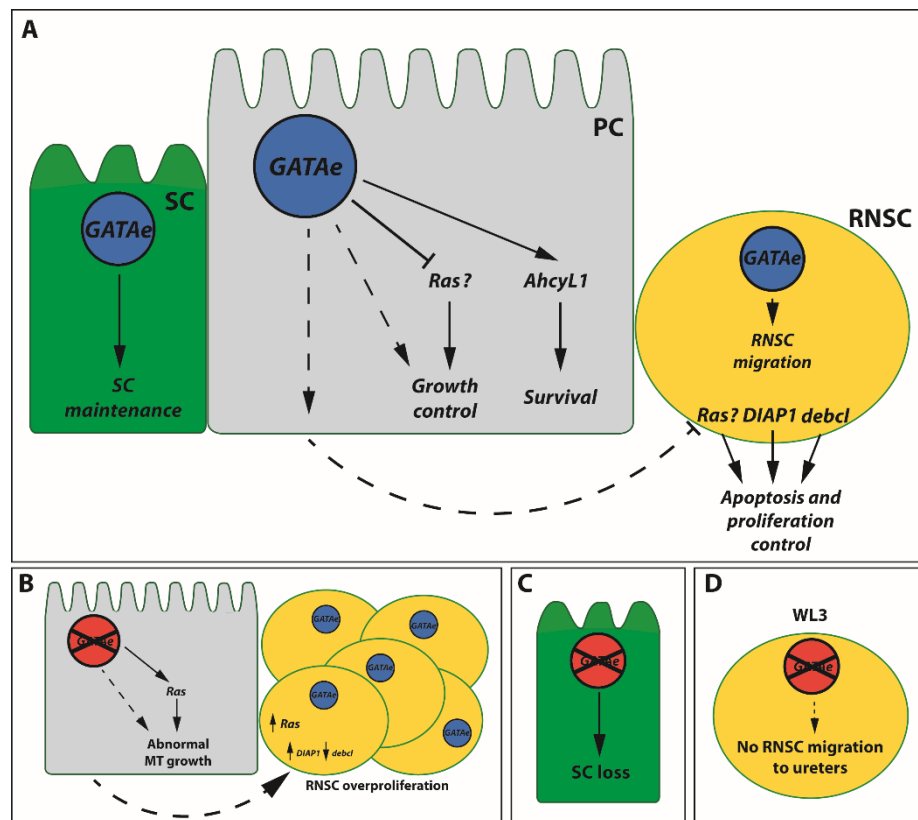


Figure 5.5.1 Proposed model for the function of *GATAe* in the MTs.

(A) Regulatory interactions of *GATAe* in the three different cell types of the MTs; PCs (grey), SCs (green) and RNSCs (yellow). Bar arrows indicate possible inhibition of gene expression. Dotted arrows indicate interaction through unknown pathways. In SCs, *GATAe* is required for their survival. In PCs, it is required for the proper architecture of the MTs possibly through Ras signalling and other unknown pathways. *GATAe* is also necessary in the PCs to ensure the normal proliferation of RNSCs through the activity of *DIAP1* and *debcl*. In parallel, *GATAe* is required in RNSCs for their migration to the ureters during metamorphosis. *Ras* is showed with a question mark since it is not fully clear if it's being modulated directly in PCs, in RNSCs, or in both cell-types. (B) In *GATAe* knockdown conditions in the PCs, the MTs show a phenotype of abnormal growth, possibly through Ras signalling and additional pathways, and induce RNSCs to overproliferate. (C) Loss of *GATAe* in SCs is associated with a reduction in this cell population. (D) *GATAe* is required in prospective RNSCs at wandering L3 (WL3) stage for their proper migration to the ureters. Loss of *GATAe* in this stage induces a complete abolition of RNSCs in the adult ureters and lower tubule.

Chapter 6 - Molecular and biochemical tools for *GATAe* functional characterisation

6.1 Introduction

The morphological and physiological implications of cell-specific *GATAe* function in the MTs have been extensively characterised and presented in the last two chapters. This chapter describes the techniques and approaches to investigate the molecular mechanism(s) of *GATAe*.

In order to precisely characterise the expression pattern and tissue localisation and possible protein-protein/protein-DNA interactions of *GATAe*, an antibody that recognises the isoform A of *GATAe* was generated.

A gain-of-function line for *GATAe* was also employed in order to examine the effects of *GATAe* overexpression. This transgenic strain contained an HA epitope tag in the C-terminal of *GATAe* coding sequence, which was also used for immunoprecipitation (IP) and Chromatin Immunoprecipitation (ChIP) experiments.

6.2 Predicted 3D structure of *GATAe* protein

Zinc fingers are one of the most abundant DBDs in eukaryotes (Klug and Rhodes, 1987). Identifying the number and structure of the zinc fingers of a TF is not only useful to study its functions, but it is also utilised to establish evolutionary links between proteins of the same family (Krishna et al., 2003). The structure of *GATAe* and the localisation of its potential zinc finger domains have been investigated in this project. Several online databases have been used to obtain a putative 3D structure of *GATAe* protein and predict its zinc finger domains. SWISS modelling analysis demonstrated a single zinc finger isoform A of *GATAe* (Figure 6.2.1A). In addition, all online databases employed in this project (see 2.10), using any of the isoforms of *GATAe* as a template (both the transcript or the resulting amino acid sequence) also resulted in only one canonical zinc finger domain (Figure 6.2.1B). The zinc finger domain detected with these methods is the most conserved region between *GATAe* and *Drosophila* or human GATA factors (Okumura et al., 2005, Lowry and Atchley, 2000).

These results contrast with the first publication about *GATAe*, as it indicated that *GATAe* possesses two zinc finger DBDs, (Okumura et al., 2005). According to

Okamura et al., the N-terminal zinc finger domain of *GATAe* (which has not been found here) is not a typical one, as it is five aa shorter compared to the C-terminal zinc finger of *GATAe* (Okumura et al., 2005). However, their findings contrast with a later report that showed that the majority of *Drosophila* GATA factors (including *GATAe*) possess only one zinc finger, instead of two (He et al., 2007).

Whereas the patterns of expression and biological functions of *GATAe* (and other invertebrate GATA factors) have been explored in detail (Okumura et al., 2005, Senger et al., 2006, Okumura et al., 2016, de Madrid and Casanova, 2018, Zhai et al., 2017, Dobson et al., 2018), the core studies of the structure and phylogeny of GATA factors and their zinc finger domains have been mostly focused on vertebrates (Lentjes et al., 2016, Morrissey et al., 1997, Lowry and Atchley, 2000). Also, the most complete characterisation of the different domains of GATA factors including *Drosophila* data was performed before the discovery of *GATAe* (Lowry and Atchley, 2000), and, in consequence, did not include any data regarding *GATAe*. The data showed in this thesis, in accordance with previous reports (He et al., 2007), indicated that *GATAe* only possesses one zinc finger domain, instead of two.

In addition, the modelling methodology used in the first study that identified two zinc finger domains in *GATAe* was not specified (Okumura et al., 2005). Therefore, it is very likely that the methodology used for zinc finger identification in that report differs with the modelling tools used in this thesis and consequently, the number of zinc finger domains detected was different. It is important to note that *srp* (another *Drosophila* GATA factor related to haematopoiesis, see 1.4.3.2) was first shown to contain a single GATA zinc finger. However, it was later demonstrated that, by alternative splicing, some isoforms of *srp* contain two zinc finger domains instead of one, conferring specificity to its functions (Waltzer et al., 2002). However, it is unlikely that this is relevant to *GATAe* since regardless of the isoform used as a template (A, B or C), analysis has always resulted in one zinc finger domain.

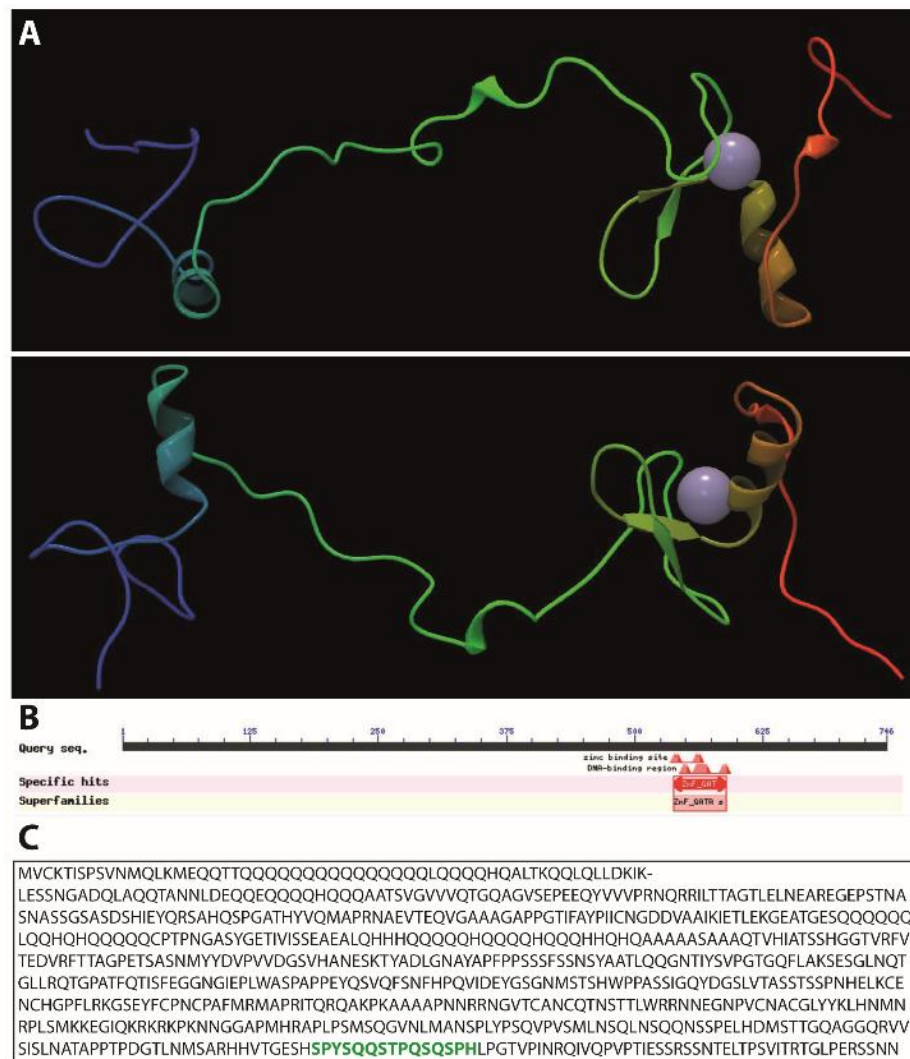


Figure 6.2.1. Predicted 3D structure, sequence, and domains of GATAe protein.

(A) Putative SWISS model of the 3D sequence of GATAe protein, viewed from two different angles. The Zn^{2+} ion is shown as a grey ball. (B) PBLAST analysis of the isoform A of GATAe. The only zinc finger domain detected is shown in red. (C) Complete GATAe amino acid sequence (isoform A). The chosen sequence for antibody design (see below) is highlighted in green.

6.3 Generation of an antibody against GATAe

6.3.1 Sequence selection

Since isoform A of GATAe is the most highly expressed GATAe isoform in the adult MTs (see 1.4.4.1), an antibody was generated using an antigenic sequence (epitope) of this specific isoform. Ab designer online software (Pisitkun et al., 2012) was utilised according to 3 parameters: (i) The Ig-score rank (which lists all the potential peptides from the most to least immunogenic sequences), (ii) Uniqueness-optimized rank (which ranks the peptides regarding their specificity to the selected protein), (iii) Conservation-optimised rank (which compares different species and scores the sequences from the most conserved to the

least). The isoform A of GATAe protein was used as a template, and a 15 amino acid **SPYSQQSTPQSQSPH** sequence was selected (Figure 6.2.1C). Although this sequence scored 7th in all three ranks, it was selected as the other candidates that showed higher immunogenicity were all located within the GATA zinc finger DBD. As the zinc finger domain is common for all the GATA factors in *Drosophila*, and, a GATAe-specific antibody cannot be designed using the zinc finger domain epitopes. Rabbit polyclonal antibodies were generated by Genosphere Biotechnologies® (Paris, France) and the IGG fraction obtained was purified as described in 2.5.

6.3.2 Immunocytochemistry and Western-Blot using GATAe antibody.

The purified IGG fraction was analysed for their specificity against GATAe using ICC and wester-blot analysis (WB). Both the IGG fraction (crude antibody) and the peptide-affinity purified antibody were tested. Wild-type, adult MTs were dissected and immunostained using these samples. GATAe antibody was tested across a wide range of concentrations (1µg/ml, 5 µg/ml, 10 µg/ml and 20µg/ml). At any of those concentrations, while nuclear staining was expected, no clear labelling was observed in the nuclei of either PCs, SCs or RNSCs, in contrast with the results obtained employing the GATAe-Gal4 strain (see 4.2.1). ICC results from the IGG fraction and the peptide-purified antibody showed diverse patterns of expression, including Endoplasmic Reticulum (Figure 6.3.1C), perinuclear (Figure 6.3.1F), or cytoplasmic (Figure 6.3.1D, E) patterns. Additionally, the generated antibody did not recognise SCs in adult or pupal stages. However, faint but potentially specific expression in the PCs of the ureter region was detected at the adult stage in some cases (Figure 6.3.1A).

In addition, embryos stained with GATAe antibody did not exhibit any specific staining in the MG or the MTs (data not shown), two tissues that express GATAe (Okumura et al., 2005).

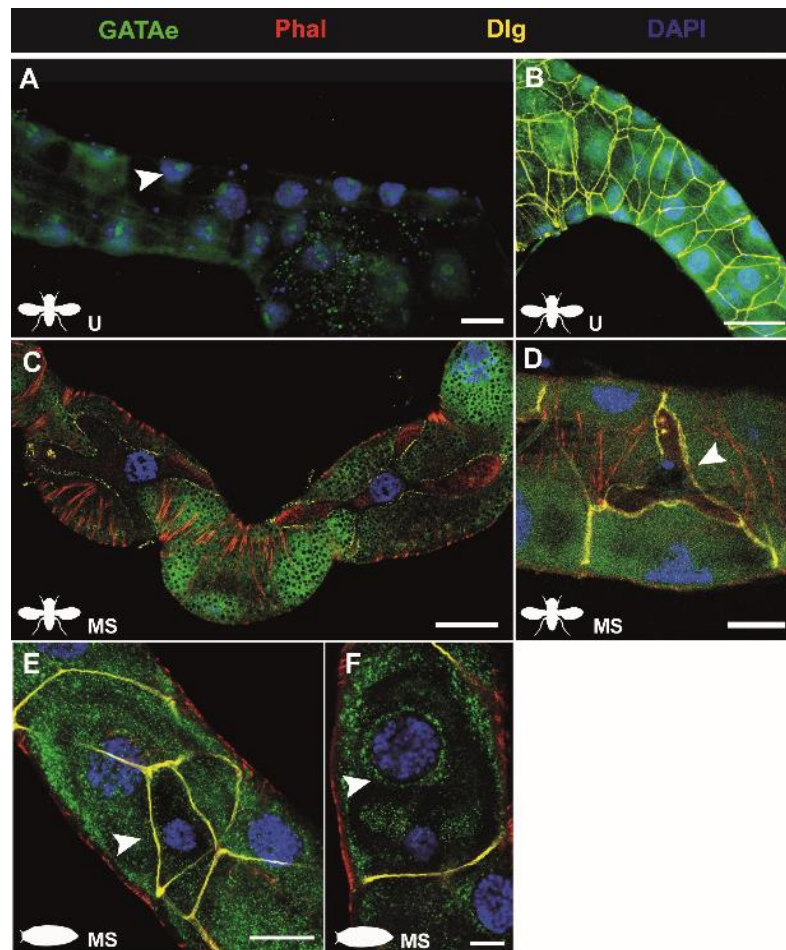


Figure 6.3.1. ICC using GATAe antibody.

ICC of adult (A-D) and pupal (E, F) MTs using the peptide-purified form of GATAe antibody. No specific staining in the nucleus was detected in either PCs, SCs or RNSCs, except faint staining in the nucleolus of the PCs in the adult stage (arrow in A). The antibody labels the ER (C), and it does not stain SCs in either adult (D) or pupal stage (E). In some cases, GATAe antibody stained a region bordering the nuclei of pupal PCs (arrow in F). U: ureter, MS: main segment. Scale bars are 50µm for C, 20µm for A, B, D and E and 10µm for F.

WB analysis was undertaken to determine the specificity of the GATAe antibody. WB analysis was initially performed from adult MTs protein samples using whole IGG fraction (non-peptide-purified GATAe antibody). Several bands could be detected, of approximate sizes of ~60, ~80 ~110kDa, and a weak band of ~160kDa (Figure 6.3.2A). Membranes were stripped and re-stained with the peptide-purified antibody, and results show that the ~60 and ~110kDa bands were detected but not the ~80 and ~160kDa (Figure 6.3.2B).

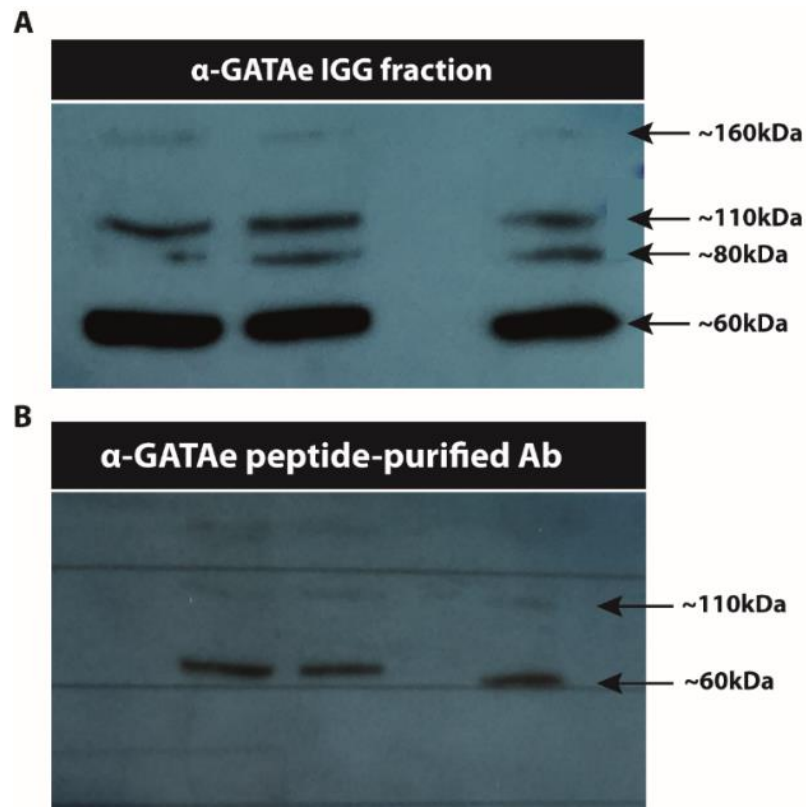


Figure 6.3.2. Western blot analysis of GATAe antibody.

(A) Analysis of three biological MT samples using the GATAe IGG fraction. (B) The same membrane of (A) was stripped and re-probed using the peptide-purified GATAe antibody.

The GATAe isoform A is 746 amino acids long, resulting in an expected band size of 80.9kDa, whereas the isoforms B and C are 731 amino acids, resulting in a protein of 79.3kDa. The results obtained from the Western blot analysis indicate that the GATAe antibody detects an 80kD protein of a similar size to GATAe, it also detects other proteins, including one at 60 kDa. Therefore, the GATAe antibody generated does not accurately detect the GATAe protein, potentially explaining the ICC data (Figure 6.3.1) and would not be suitable for subsequent ChIP-seq experiments.

6.4 Generation of a gain-of-function line of GATAe

A GATAe gain-of-function strain was generated to investigate the potential effects of GATAe overexpression. Alternative ways to induce GATAe overexpression lines have been previously created, such as the introduction of a UAS-GATAe plasmid by germline transformation, for the study of developing and adult MG (Zhai et al., 2017, Okumura et al., 2005). Here, a transgenic line which expresses a fusion protein of GATAe isoform A and an HA tag was generated (GATAe-HA, see vector map and other details in 2.2.1). This HA tag should allow

GATAe protein pull-down experiments since it was not possible to produce a GATAe-specific antibody.

The UAS-GATAe-HA construct obtained from DGRC is 11244bp long. To confirm the sequence of this plasmid, restriction enzyme digestion was carried out using different enzymes with distinct restriction patterns: KpnI and NheI, BamHI and NheI. Analysis of resulting bands (Figure 6.4.1) showed that bands of expected sizes were produced: KpnI and NheI each cut at a single site, resulting in two bands; BamHI cuts at two sites, and a combination of BamHI and NheI cuts at three sites (all cutting sites are shown in Figure 6.4.1). Additionally, bands were extracted and sequenced using specific primers against *GATAe* (data not shown). Altogether, these results confirmed the integrity sequence of the *GATAe* plasmid.

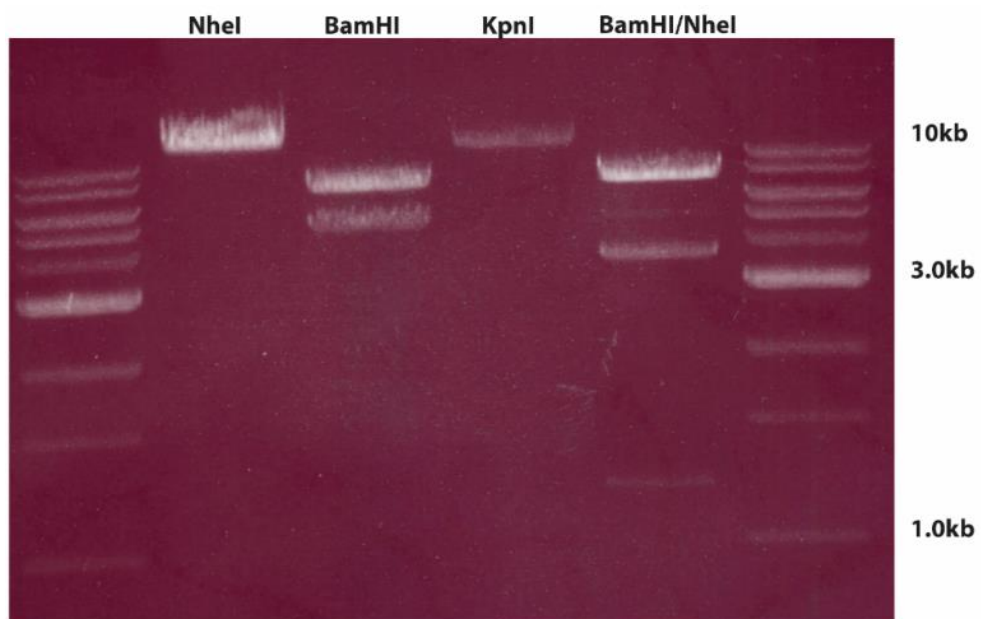


Figure 6.4.1. Restriction enzyme digest analysis of UAS-GATAe-HA plasmid.

UAS-GATAe-HA plasmid DNA was digested using restriction enzymes. A single cut with NheI or KpnI (at positions 1304 and 102, respectively) results in a linearised plasmid of a size of ~11kb. BamHI cuts the DNA twice (at 104 and 4717), resulting in fragments of 4643 and 6601bp. Digestion of BamHI and NheI cuts the DNA three times (at 104, 4717 and 1304), resulting in fragments of 1200, 3443 and 6601bp. In this lane, there is presence of a very weak band of ~4.5kb, that corresponds to partial digestion of BamHI.

6.4.1 S2 cell analysis

The UAS-GATAe-HA plasmid was transiently co-transfected together with the pMT-Gal4 plasmid in *Drosophila* S2 cells to confirm GATAe-HA protein

localisation. Results from ICC experiments show that as expected, GATAe-HA protein was exclusively localised in the nuclei of S2 cells (Figure 6.4.2).

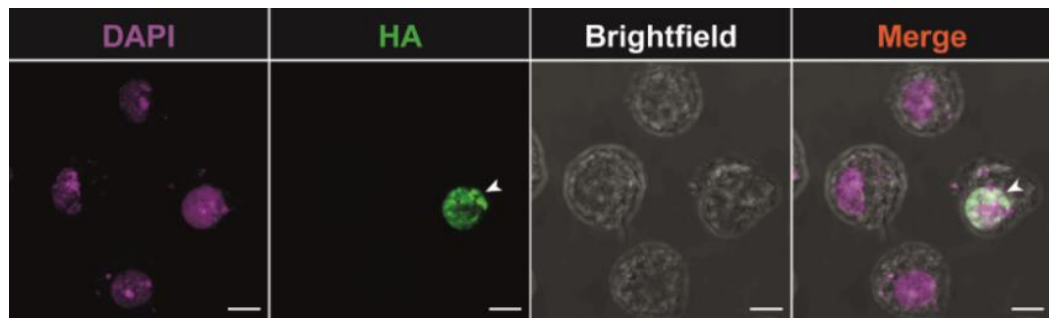


Figure 6.4.2. Nuclear localisation of GATAe-HA in *Drosophila* S2 cells.

ICC of *Drosophila* S2 cells, stained with DAPI (purple) and anti-HA (green) Only co-transfected S2 cells (white arrow) show evident nuclear HA staining, compared to untransfected cells, which only display DAPI staining (purple) The right panel shows merged channels of DAPI, HA and brightfield. Scale bars are 5µm.

Furthermore, qPCR experiments showed a significant upregulation of the *GATAe* gene expression (>25 fold) in the *GATAe*-HA transfected cells compared to untransfected cells (Figure 6.4.3). Previous research has shown that overexpression *GATAe* induces the expression of *inx7* and *intB* two genes downstream of *GATAe* in the developing MG (Okumura et al., 2005, de Madrid and Casanova, 2018). Therefore, *inx7* and *intB* expression were measured in *GATAe*-HA transfected S2 cells. As expected, *GATAe*-HA transfected S2 cells exhibited significant upregulation of expression of both *inx7* and *intB* (~5 and ~4-fold increase, respectively), demonstrating that the UAS-*GATAe*-HA plasmid is effective as an overexpression construct with no detrimental impact of the HA tag.

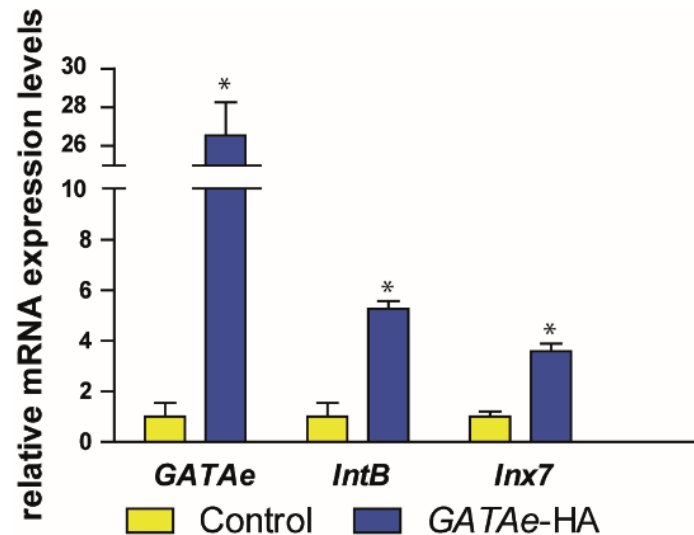


Figure 6.4.3. *GATAe* and putative downstream gene expression in S2 cells transfected with UAS-*GATAe*-HA.

Relative mRNA levels, comparing S2 cells transfected with *GATAe*-HA (blue bars) and control S2 cells (untransfected, yellow bars). Compared to the control, *GATAe*-HA co-transfected S2 cells show strong expression levels of *GATAe*, as well as the downstream factors *intB* or *Inx7*. Relative mRNA levels were normalised to the *rpL32* expression. The bars marked with * indicate significant difference ($p < 0.05$, Student T-test, two-tailed).

6.4.2 Tissue-specific overexpression in *Drosophila*

Several independent transgenic fly lines were generated commercially in this study, using the UAS-*GATAe*-HA plasmid (see 2.2.1). These lines were used to overexpress *GATAe* in a cell-specific manner using PC and SC GAL4 drivers. GAL4 drivers for ubiquitous expression were also used for comparative studies with the PC and SC-specific lines. Results from ICC experiments using *CapaR*>UAS-*GATAe*-HA adult MTs showed that the pattern of expression of the *GATAe*-HA protein was limited to the nuclei of PCs, in line with the S2 cell data (Figure 6.4.4A-E). Accordingly, qPCR experiments confirmed that *GATAe* mRNA levels were significantly increased in *CapaR*>UAS-*GATAe*-HA MTs compared to the control (Figure 6.4.4H). Interestingly, *CapaR*>UAS-*GATAe*-HA MTs did not show any severe morphological malformations. However, compared to the controls, the *CapaR*>UAS-*GATAe*-HA MTs exhibited abnormally big aggregations in the lumen of the initial segment (Figure 6.4.4F and G), resembling kidney stones (Chi et al., 2015). Similar to the S2 cells, adult *CapaR*>*GATAe* HA MTs also showed increased expression of *GATAe* downstream genes *IntB* (~1.7-fold increase), and *Inx7* (~3-fold increase).

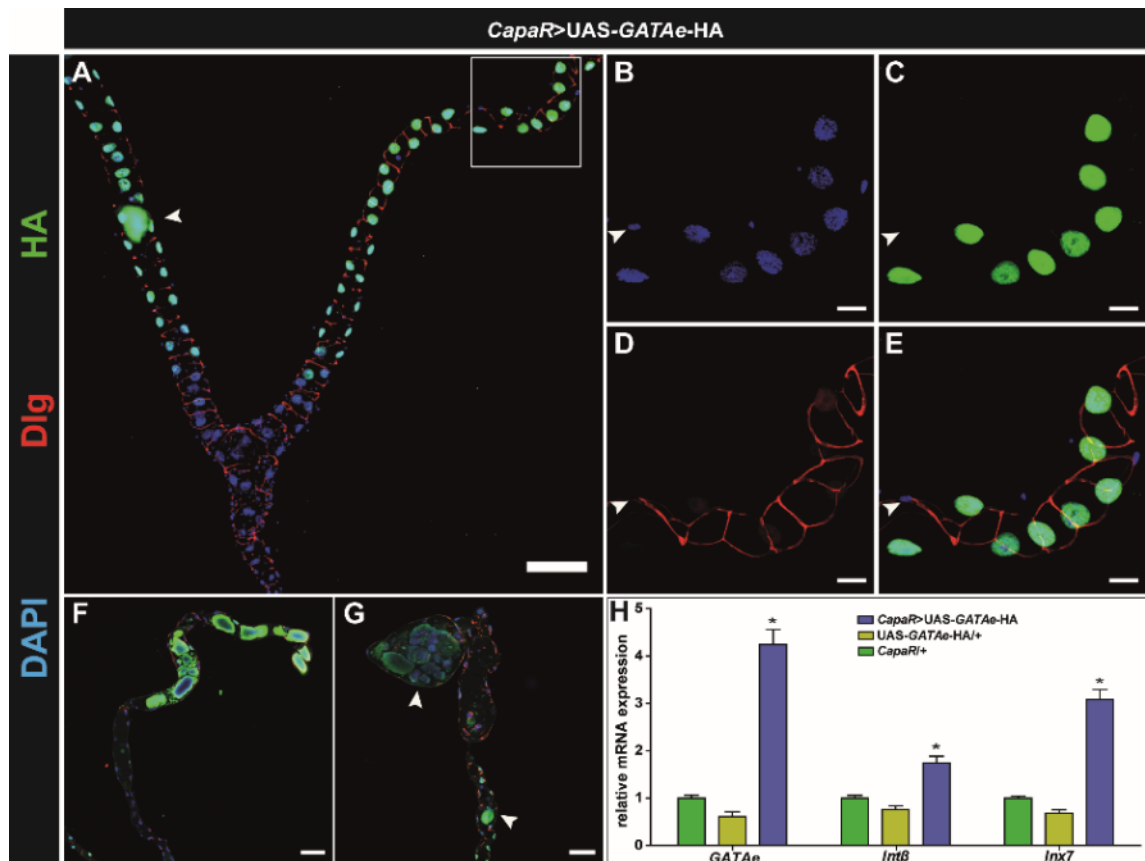


Figure 6.4.4. Characterisation of *CapaR*>UAS-*GATAe*-HA adult MTs.

(A) Expression of *GATAe*-HA via *CapaR*-Gal4 localises *GATAe*-HA in the nuclei of PCs. (B-E) Magnification of the area highlighted in A, showing exclusion of expression in SC nuclei (white arrow in B-E). (F and G) Examples of concretions produced in the initial segment of adult tubules overexpressing *GATAe*. Most of them are localised in the initial segment (green in F-G), but they can also be found in other regions of the tubule (white arrows in A and G). (H) Relative mRNA quantifications of *CapaR*>UAS-*GATAe*-HA (blue bars) and both control (green and yellow bars) adult MTs. *CapaR*>UAS-*GATAe*-HA MTs show significantly higher levels of *GATAe*, (~4-fold increase) and both its targets, *intβ* and *inx7* (~1.7 and 3-fold increase respectively). Relative mRNA expression levels were normalised to the *rpl32* expression. ($p < 0.05$, two-tailed student T-test). Scale bars are 100 μ m in A, F and G, and 20 μ m in B-E.

UAS-*GATAe*-HA expression was also driven with *CIC-a*-Gal4 to observe possible phenotypes in SCs, and driven with *ctB*-Gal4 to examine possible defects in embryonic development of the MTs. *CIC-a*>UAS-*GATAe*-HA MTs did not show any detectable defects (Figure 6.4.5A). In contrast, *ctB*>UAS-*GATAe*-HA and ubiquitous *Actin*>UAS-*GATAe*-HA larvae died before reaching pupariation, probably during in L1/L2 stages, as no WL3 specimens were observed. In *ctB*>UAS-*GATAe*-HA embryos, HA staining colocalised almost exclusively with the PC marker *ct*, indicating that HA expression was restricted to the PCs (arrows in Figure 6.4.5B). However, in other regions, such as the developing nervous system, HA staining was not detected, and only *ct* staining was observed, suggesting that *ctB*-Gal4 expression is limited to the developing MTs, in contrast to the extended pattern of *ct* antibody staining. It has been already reported

that *ctB* is expressed only in a subset of the native *ct* expression pattern (Jack and DeLotto, 1995). Interestingly no visible morphological phenotypes were observed in MTs of these embryos, so causes of larval lethality remain to be explored.

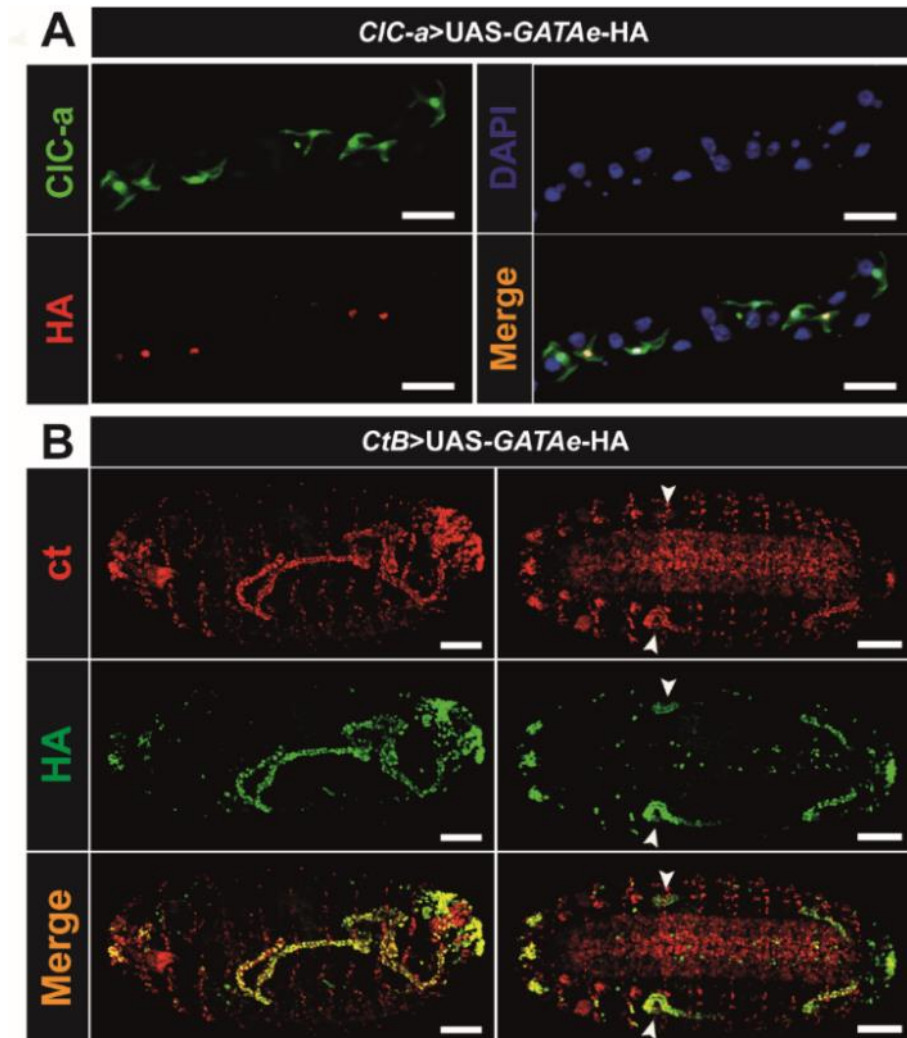


Figure 6.4.5. Overexpression of UAS-GATAe-HA using different Gal4 lines.

(A) Adult *CIC-a>UAS-GATAe-HA* MTs stained using antibodies against *CIC-a* (green), HA (red) and DAPI (blue). *CIC-a>UAS-GATAe-HA* MTs show evident HA staining in the nucleus of SCs. (B) Two examples of *ctB>UAS-GATAe-HA* embryos stained against *ct* (red) and HA (green). Both show a robust colocalization in developing MTs, but HA is not detected in the midline of the developing ventral nerve cord. Scale bars are 50µm.

Rescue of the morphological phenotypes induced by *GATAe* silencing (see 4.2.2) was attempted by combining the constructs *GATAe* RNAi and UAS-*GATAe-HA* (*CapaR>GATAe RNAi;UAS-GATAe-HA*). Unfortunately, *CapaR>GATAe RNAi;UAS-GATAe-HA* MTs phenocopied *CapaR>GATAe RNAi* MTs, suggesting that the effects of RNAi silencing of *GATAe* are epistatic over *GATAe* overexpression effects (as the RNAi line also targets the mRNA produced from the *GATAe-HA* construct). In other words, that the high levels of *GATAe* induced by UAS-*GATAe-HA* are not

able to compensate for the low levels of *GATAe* caused by RNAi silencing. A possible explanation of this issue is discussed in 6.6.1.

6.5 Chromatin Immunoprecipitation Analysis of *GATAe*

6.5.1 Chromatin Immunoprecipitation

Chromatin Immunoprecipitation (ChIP) is a potent technique to identify specific DNA binding regions for particular TFs, allowing the study and characterisation of signalling pathways and genetic interactions (Ghavi-Helm et al., 2016). In brief, a DNA-binding protein, such as a TF, is fixed and immunoprecipitated, and any DNA sequences bound to the protein of interest can be identified by sequencing techniques (Chromatin Immunoprecipitation Sequencing, or ChIP-Seq). ChIP-Seq has not only been used for the study of TFs, but it is also widely used to localise post-translationally modified histones, histone variants or enzymes associated with chromatin (Collas, 2010, Shoaib, 2011). This next chapter focused on the study of *GATAe* using ChIP-Seq and highlights troubleshooting that occurred during the optimisation of this technique.

6.5.2 Technical ChIP optimisations

There are several considerations when designing ChIP-seq experiments. Here, the first aim was to determine the proper quantity of tissue which ultimately yielded sufficient DNA concentration for sequencing. For ChIP-seq experiments with *Drosophila*, 150 imaginal wing discs (Oh et al., 2013), 10-50 millions of *Drosophila* S2 cells (Gilchrist et al., 2009), or 100mg of whole embryos (Menoret et al., 2013) were previously used. The commercially-based ChIP protocol used in this project (see 2.8) recommended an amount of 25mg of starting tissue in order to recover good DNA yields for sequencing. This amount corresponds to approximately 4000 dissected adult MTs (~1000 adult flies) for each sample. Therefore, for optimisation of the ChIP protocol before MT dissections, initial ChIP-Seq experiments were performed using S2 cells overexpressing the *GATAe*-HA construct.

6.5.3 Results

6.5.4 Immunoprecipitation analysis

To determine if the antibody-coupled magnetic beads can specifically bind to the HA tag, both S2 cell samples (co-transfected with pMT-Gal4 and UAS-GATAe-HA plasmids) and adult dissected MT samples overexpressing GATAe-HA (*CapaR*>UAS-GATAe-HA) were immunoprecipitated and analysed by WB. WB analysis against the HA tag showed a strong band of 110 kDa in both co-transfected S2 cells and *CapaR*>UAS-GATAe-HA immunoprecipitated fractions (Figure 6.5.1A1 and B1), whereas no such band was detected in negative controls (untransfected S2 cells and UAS-GATAe-HA adult MTs, Figure 6.5.1A3 and B2 and 3). Additionally, a band of around 50kDa was also detected in all IP experiments, which most probably corresponds to the heavy chain of the HA-antibody that is expected to co-elute with the IP sample (Janeway Jr et al., 2001). Although the detected band of 110 kDa is unique and corresponds to the GATAe protein, it differs from the theoretical molecular weight of GATAe (80kDa). This would be due to possible post-translational modifications of the protein, a common feature for TF regulation (Tootle and Rebay, 2005, Bode and Dong, 2004), which is also present in GATA TFs (Bresnick et al., 2012, Menghini et al., 2005). Possible post-translational modifications of the GATAe protein would explain the many different functions that this TF can have depending on the cellular context. Research to highlight these modifications and their significance would be an exciting direction for future studies. As expected, this specific 110 kDa band is not detected in samples that have not been subjected to IP or in the negative controls (see legend of Figure 6.5.1).

Altogether, these results, demonstrate that the GATAe-HA protein is specifically immunoprecipitated using an anti-HA tag antibody and which therefore permits subsequent ChIP-seq experiments.

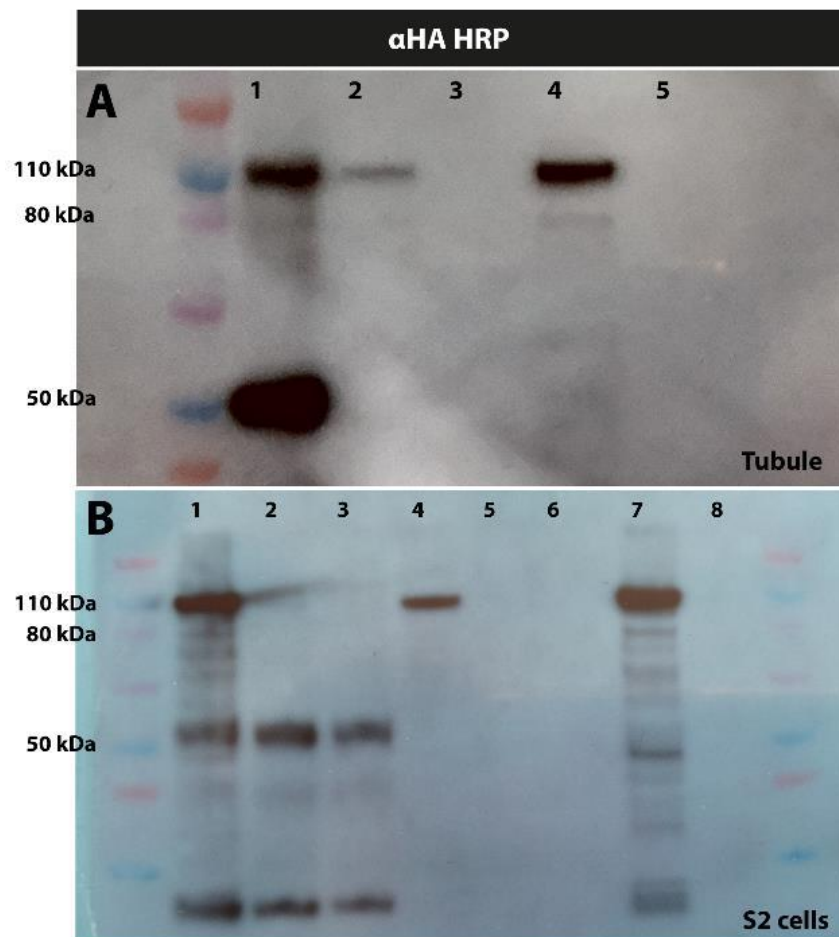


Figure 6.5.1. Western-blot analysis of IP experiments for HA tag.

(A) IP of adult *CapaR>UAS-GATAe-HA* MTs (lane 1) or parental control line *UAS-GATAe-HA/+* MTs (lane 3) using anti-HA antibody. A strong band of 110 kDa is detected only when *GATAe-HA* is overexpressed (lane 1). Lane 2 corresponds to the first wash after bead incubation. A weak band of 110 kDa is detected indicating that only a small amount of HA-tagged protein is lost during the first wash. Lanes 4 and 5 are positive (dissected adult *CapaR>UAS-GATAe-HA* MTs) and negative (dissected adult *UAS-GATAe-HA/+* MTs) controls, which have not been subjected to IP (B) IP of S2 cells which have been co-transfected with *Des-Gal4* and *UAS-GATAe-HA* plasmid (lane 1) or two independent negative control samples (untransfected S2 cells, lanes 2 and 3). Lanes 4, 5 and 6 correspond to the first wash after the incubation with the magnetic beads, of the IP samples from the lanes 1, 2 and 3, respectively (lane 4, which corresponds to the co-transfected sample, also shows a weak band of 110kDa, whereas lanes 5 and 6 are empty). Lanes 7 and 8 are positive (co-transfected *UAS-GATAe-HA* S2 cells, lane 7) and negative controls (untransfected S2 cells, lane 8), which have not been subjected to IP.

6.5.5 Chromatin Immunoprecipitation

6.5.5.1 S2 cells

ChIP experiments were performed on the basis that the protocol for IP of *GATAe-HA* had been optimised. For ChIP, as previously mentioned, S2 cells were initially used to optimise the protocol. 1×10^7 cells were used for each ChIP sample. The commercial kit employed in this study recommended 4×10^6 mammalian cells (see 2.8.1.1). Given the smaller size of S2 cells compared to mammalian cells, in this study 1×10^7 S2 cells were used. Cells were crosslinked

and treated with Micrococcal nuclease (Mnase) for 20 min. Mnase treatment results in the fragmentation of the DNA between 150bp and 900bp, equivalent to 1 to 5 nucleosomes, which is the optimal size for sequencing (Figure 6.5.2). After Mnase treatment, and before IP, samples were run on an agarose gel, and chromatin concentration was calculated according to the manufacturer's protocol. The DNA concentration in all chromatin samples was calculated as ~50 ng/ μ l. However, despite following the standard protocol, after performing ChIP with the recommended antibody concentration (5 μ g of antibody for each assay), the final DNA concentrations in both IPd and non-IPd samples were ~0.12ng/ μ l, which suggests that *GATAe* HA-specific DNA was not present.

Subsequently, several attempts were made to optimise the protocol and obtain significant ChIP DNA from S2 cells as follows: (i) concentration of anti-HA antibody for the pulldown was increased from 5 μ g up to 10 μ g for each assay. (ii) antibody incubation times were increased from 2h to overnight. (iii) time of protein elution from the beads was increased from 30min to overnight where bead elution and DNA de-crosslinking were included in the same step. (iv) Final DNA was eluted using a lower volume of elution buffer (30 μ l) or via a MinElute® QIAGEN DNA extraction kit which elutes DNA in 10 μ l, in order to concentrate DNA in the eluant fraction. However, none of these protocol optimisations was sufficient to obtain significant DNA concentrations from the chromatin preparation.

Possible explanations for the lack of DNA include: (i) Transfection efficiencies in S2 cells are not sufficiently efficient to induce high expression of *GATAe* protein and subsequent DNA pulldown (the exact transfection efficiency levels are unknown, but as two plasmids needed to be transfected, they are presumably low). (ii) DNA degraded, as nuclease treatment could be too intense.

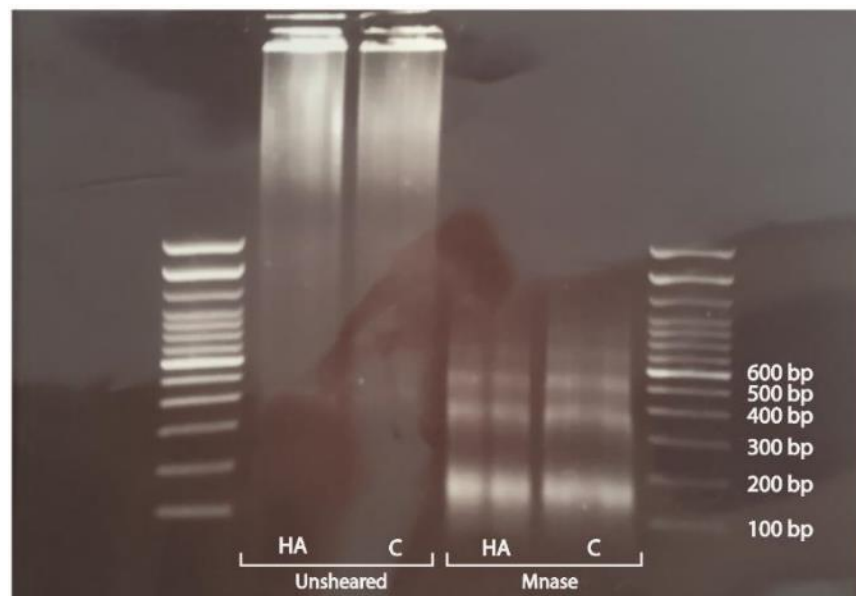


Figure 6.5.2. Micrococcal nuclease treatments of S2 cells.

DNA of S2 cells was decrosslinked and purified before Mnase treatment (unsheared, left two lanes) and after 20min of Mnase treatment (Mnase, right two lanes). After nuclease digestion, DNA is cut in small fragments, majorly between 150bp and 900bp. HA indicates the DNA extracted from S2 cells co-transfected with *Des-Gal4* and *UAS-GATAei-HA*, and C indicates the DNA extracted from untransfected S2 cells (negative control).

To determine which of these two hypotheses is correct, an additional ChIP experiment was performed using a high cell number ($\sim 7 \times 10^7$) cells for each condition, amounting to $\sim 20\times$ the number of cells required for each ChIP in the original protocol. This may probably compensate for the S2 transfection rates. In this case, a detectable quantity of DNA was measured in the pulldown pMT-Gal4>UAS-*GATAe*-HA S2 sample. However, non-transfected samples had higher DNA concentrations. (Table 6.5-1A and B). Furthermore, the final DNA concentrations were extremely low.

Previous reports have shown that low resulting DNA concentrations could be caused by poor cell lysis and therefore low DNA recovery (Bortz and Wamhoff, 2011). To check this eventuality, pellets from step 2.8.2 of the ChIP protocol were resuspended and subjected to harsher sonication cell treatment.

Interestingly, the final DNA concentration in the IP fraction using this method was drastically improved. However, the untransfected samples also showed higher final DNA concentrations compared to *Des*>UAS-*GATAe*-HA S2 cells (Table 6.5-1C and D).

Sample	DNA concentration (µg/ml)
(A) <i>Des-Gal4>UAS-GATAe-HA</i> (S2 cells)	5
(B) Untransfected (S2 cells)	20
(C) <i>Des-Gal4>UAS-GATAe-HA</i> (Cell pellets)	68
(D) Untransfected (Cell pellets)	720

Table 6.5-1. Final ChIP DNA concentrations of S2 cells

ChIP DNA concentrations of *Des-Gal4>UAS-GATAe-HA* transfected S2 cells (A and C) and non-transfected S2 cells (controls, B and D). In all cases, DNA concentrations in the non-transfected controls are increased compared to *Des-Gal4>UAS-GATAe-HA* S2 cells.

6.5.5.2 Adult whole-fly and MTs

As optimisation of the ChIP protocol in S2 cells failed to yield GATAe-HA-specific DNA, the ChIP procedure was attempted in whole-fly samples ahead of MT preparations. Overexpressing GATAe-HA ubiquitously using *Actin-Gal4* resulted in embryonic lethality, so viable *CapaR>UAS-GATAe-HA* adult flies were used instead. *CapaR* is mainly and strongly expressed in the MTs (Leader et al., 2018), but interestingly, it is also expressed in some regions of larval and adult MG musculature (Terhzaz et al., unpublished). For each ChIP, 50 *CapaR>UAS-GATAe-HA* adult flies (~50g of total tissue) were used, but after carefully following the optimised protocol used for S2 cells, no ChIP DNA was detected in any of the adult fly samples.

Despite this, MT samples from 1000 adult female flies (2000 MT pairs, (from *CapaR>UAS-GATAe-HA* and *UAS-GATAe-HA/+*) were dissected and subjected to the optimised ChIP protocol in order to identify possible downstream genes regulated by GATAe in the PCs. Pellets from step 2.8.1.3 were collected and sonicated to retrieve the optimal amount of DNA. Again, a second attempt with a higher number of sonication cycles was necessary and finally resulted in detectable DNA after ChIP. However, like the S2 cell samples, the control sample (*UAS-GATAe-HA/+*) resulted in higher DNA concentration compared to *CapaR>UAS-GATAe-HA*.

Sample	DNA concentration (µg/ml)
(A) <i>CapaR</i> >UAS- <i>GATAe</i> -HA	Not detected
(B) UAS- <i>GATAe</i> -HA/+	Not detected
(C) <i>CapaR</i> >UAS- <i>GATAe</i> -HA	3
(D) UAS- <i>GATAe</i> -HA/+	6

Table 6.5-2. ChIP DNA concentrations of adult MTs

ChIP DNA concentrations of *CapaR*>UAS-*GATAe*-HA (A and C) and UAS-*GATAe*-HA/+ adult MTs (negative controls, B and D). C and D are the pellets collected in step X from A and B, respectively, which have been then subjected to stronger sonication cycles. ChIP DNA was only detected in C and D. Also, DNA concentrations in the negative controls are increased compared to *Des-Gal4*>UAS-*GATAe*-HA S2 cells.

Although DNA concentrations in both S2 cell and MTs samples were higher in the control samples, all samples in which ChIP DNA was detected were further analysed at Glasgow Polyomics and subjected to ChIP-sequencing.

6.5.5.3 ChIP-Sequencing analysis

Unfortunately, S2 cell samples were not able to be further processed due to the large size of the DNA fragments. However, two MT samples (*GATAe*-HA pulldown and negative control, C and D in Table 6.5-2, respectively), were fragmented more efficiently and were processed for sequencing analysis. After library preparation, quality control and sequencing, peak annotation was performed by Glasgow Polyomics facilities. The resulting peaks and gene lists were further analysed as described in 2.8.6. It is extremely important to note that, although the data showed in the next section is interesting, it must be considered as preliminary, as only one replicate is included. Future ChIP-Seq replicates, and further research will be required for validation of these data (see 6.6.1).

6.5.5.4 Motif analysis

Collaborators from Glasgow Polyomics analysed the significant peaks detected using two different False Discovery Rates (FDR, (Benjamini et al., 2006)). On the one hand, using an FDR of 0.1, more peaks were detected, but the reliability of the analysis could be potentially low. On the other hand, using an FDR of 0.01 should result in fewer but more reliable peaks. Therefore, only the peaks detected using an FDR of 0.01 were further analysed in this thesis, to minimise false positives. Using two sets of parameters for different histone modifications (H3K4me3 and H3K27me3, see 2.8.6) a total of 108 differently expressed peaks (comparing *GATAe*-HA bound chromatin vs control chromatin) were detected (for

the complete list of genes associated to these peaks, Appendix 10). Furthermore, the services from Glasgow Polyomics also performed motif identification using HOMER (2.8.6 and (Heinz et al., 2010)). Only three different known motifs were significantly enriched in the peaks found, and those were two promoter motifs and a GATA motif (Figure 6.5.3). Thus, these data reinforce the roles of *GATAe* as a TF, binding to promoter regions to regulate transcription, possibly by binding to the GATA motifs of downstream genes.

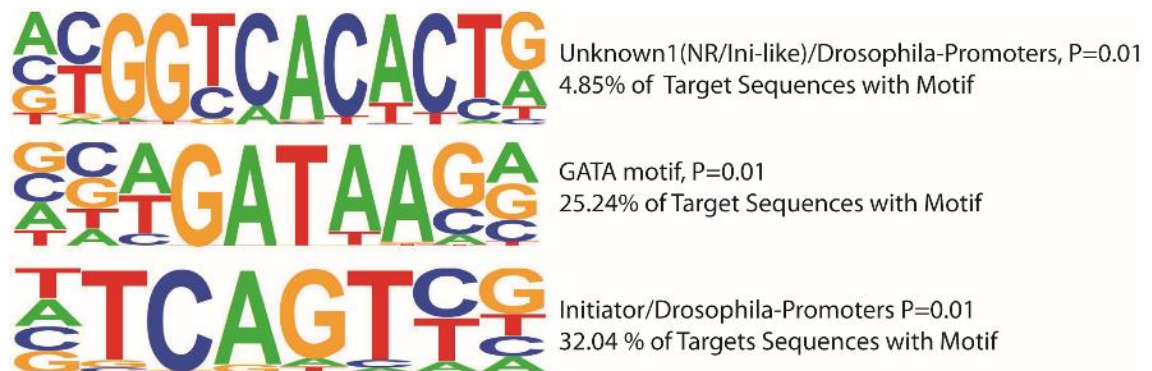


Figure 6.5.3. Enriched known motifs identified 1kb of the surrounding peaks detected.

Significantly enriched ($P=0.01$) known motifs detected in a surrounding region of 1kb of the peak summits, using HOMER. Only three motifs were detected: Two motifs for *Drosophila* promoters (Unknown1 (NR/Ini-like), present in 4.85% of Target sequences and Initiator, present in 32.04% of Target Sequences) and a GATA factor motif, present in 25.24% of Target Sequences.

Also, several *de novo* enriched motifs were identified, which are shown in Figure 6.5.4. Peaks specifically bound by *GATAe*-HA exhibited overrepresentation of the transcriptional initiator TATA box. Interestingly, the motif GATA15 was also enriched in the peaks, in accordance with Figure 6.5.3. Also, known motifs for developmental regulators were enriched in the *GATAe*-HA bound peaks, such as the homeobox bagpipe (*bap*, CTTAAGAT, 10.68% of targets, (Furlong et al., 2001, Azpiazu and Frasch, 1993)), or the fork-head family TF sloppy paired 1 (*slp1*, TTTTAAATAT, 15.57% of targets). Furthermore, motifs for tailless (*tll*, AAGTCT, 3.88% of targets), a master regulator for the development of the hindgut (which, in the embryonic development of the gut, activates the expression of *brachyenteron*, a hindgut-specific gene repressed by *GATAe* (Okumura et al., 2005, Kispert et al., 1994)) have been detected in the peaks. It would be possible that in the adult MTs, the interaction between *GATAe* and another TF with a DBD domain similar to *tll* (as *tll* does not seem to be expressed in the adult MTs, according to FlyAtlas2). Possible candidates would be *HNF4* or *Hr39*, two TFs with a highly similar DBD to *tll* (Maglich et al., 2001, Fahrbach et al., 2012), which are strongly expressed in the adult MTs. It is important to note

that the *de novo* motifs found are displayed by HOMER according to the closest known match and, in some cases could not correspond to the known motifs (Heinz et al., 2010). For instance, the enriched motif AT1G20910 found by HOMER (15.53% of targets, Figure 6.5.4), is also highly matching with the tll motif, as it has as well a strong alignment similarity.

Altogether, the known and *de novo* motifs found in this first replicate of the ChIP-Seq are a useful start to identify interactors of *GATAe* in the adult PCs. Therefore, performing at least other four extra replicates would profoundly improve the significance of the detected peaks, and provide valuable novel information.

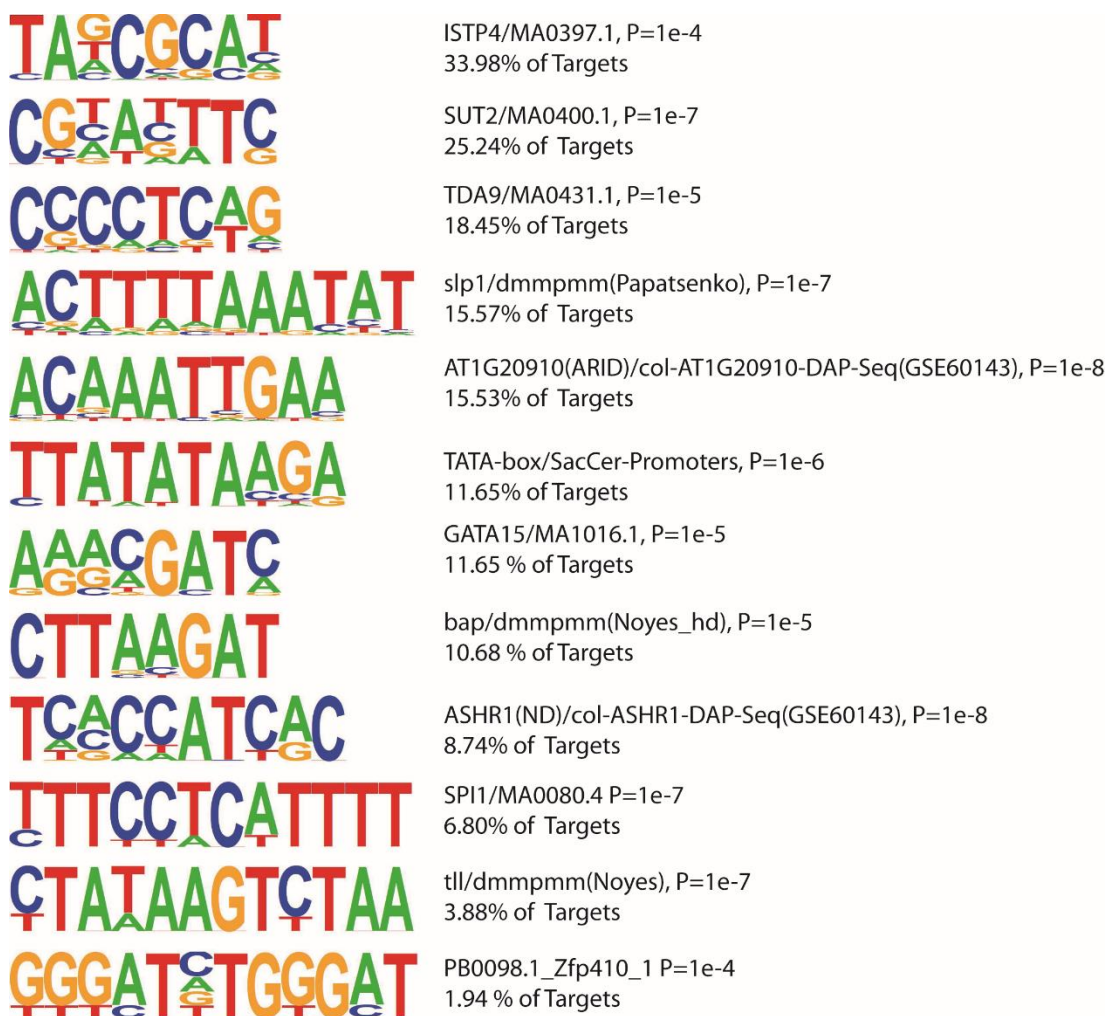


Figure 6.5.4. *De novo* motifs enrichment results

De novo motifs enriched in the *GATAe*-HA bound sequences, sorted by their abundance. Their closest known match, P-value, and % of hits with the motif are indicated on the right.

6.5.5.5 Gene Ontology classification

In addition, a list of genes was obtained from all the peaks, by identifying the genes neighbouring the peak regions (generated by Glasgow Polyomics), thus obtaining a list of potential direct downstream targets of *GATAe* (Table 7.3-3). This list of genes was further analysed and classified using two independent online GO classification tools: PANTHER, (Mi et al., 2013)) *GOTermMapper* (Consortium, 2004), as described in section 2.10. Interestingly, almost half of the genes (49 out of 108) are still named by their CG number, indicating that some of their functions are yet unknown and that some of them could potentially have novel roles in the MTs. Also, this illustrates the possible “phenotype gap” that exists in the *Drosophila* MTs (Dow, 2007, Dow, 2003), and highlights the importance of studying the function of novel genes.

Both PANTHER and *GOTermMapper* use different algorithms and curate different databases, including previous literature for gene classification, and as a result, the resulting clusters have different names depending on the software used. However, in broad terms, both PANTHER and *GOTermMapper* classified the genes in similar groups. Note that all hits analysed by both tools (PANTHER and *GOTermMapper*) are classified in each category independently. Therefore, as some of the genes may be involved in different processes, it is possible that one same gene could appear in two or more categories according to its different involvement on biological processes or molecular functions (e.g. one of the top hits of the analysis, the gene *longitudinals lacking*, or *lola*, appears both in “Anatomical development” and “nitrogen metabolism” in Figure 6.5.5B).

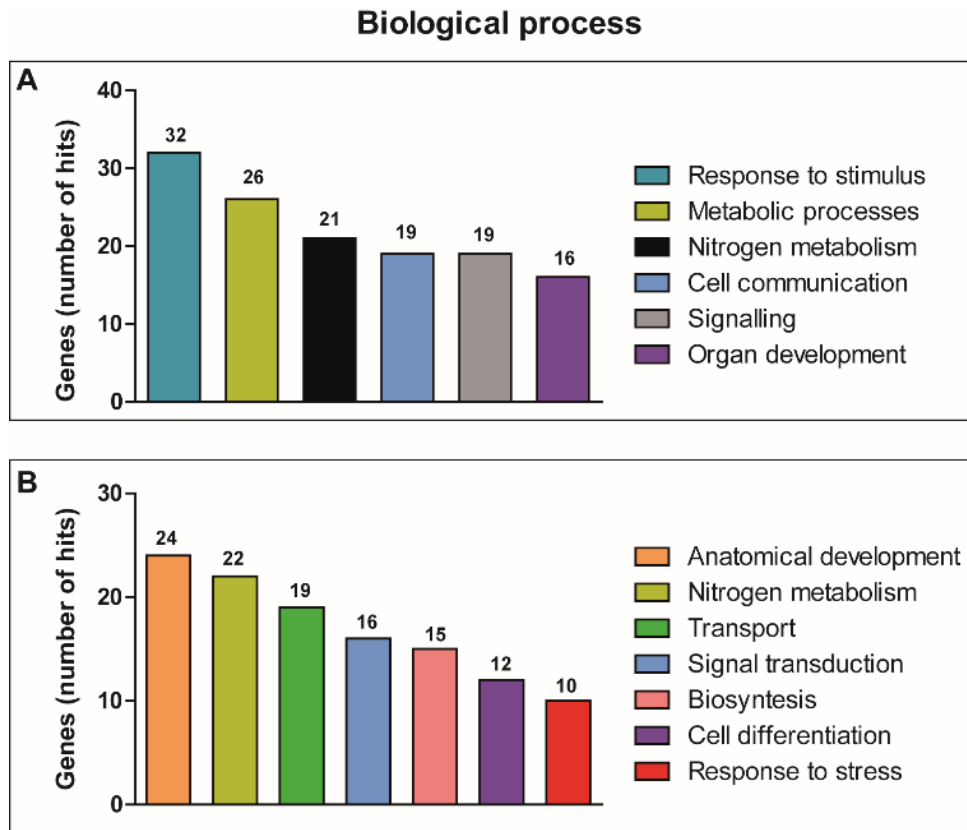


Figure 6.5.5. Classification of ChIP-Seq hits according to their biological processes.

Bar graphs representing the number of genes (hits) classified according to their biological processes, using (A) PANTHER or (B) *GOTermMapper* GO identification tools. The number of genes that correspond to each category is indicated above every bar.

According to their involvement in biological processes, most of the genes in the list were associated to metabolic processes (Figure 6.5.5A). As the term, “metabolic processes” is highly general, the 26 hits under this category (Figure 6.5.5A) were classified in more specific metabolic processes using PANTHER. Among them, 22 were associated with primary metabolic processes, 16 to macromolecule metabolic processes, 13 to organonitrogen metabolism, 8 to protein metabolism, and 6 to nucleic acid metabolism. The presence of genes involved in nitrogen metabolism in both GO analysis tools (22 genes involved in nitrogen metabolism according to *GOTermMapper*, Figure 6.5.5B) is particularly interesting, given the vital functions of the MTs in the secretion and elimination of nitrogenous wastes (Denholm and Skaer, 2009, Browne and O'Donnell, 2013). Interestingly, previous research demonstrated that GATA factors regulate nitrogen metabolism in yeast (Cooper, 2002), which could suggest an evolutionarily conserved function of GATA factors.

Also, a significant proportion of the genes are involved in response to stimulus (GO:0051869, 32 genes, Figure 6.5.5A). Again “response to stimulus” is a broad

term, the hits under the category of “response to stimulus” were as well subdivided into more specific categories. The main categories representing these hits (according to PANTHER) were “cellular” response to stimulus (22 hits), “response to external stimulus” (GO:0009581, 7 hits), response to organic substances (6 hits). Importantly, there was presence of hits involved in response to stress (GO:0006950, 10 genes, which coincided with the *GoTermMapper* data in Figure 6.5.5B). Again, this correlates with the previously reported roles of the MTs in stress responses (Davies et al., 2012, Davies et al., 2014a, Terhzaz et al., 2015b), and together with the experimental data showed in this thesis (see 4.3), suggests a role for *GATAe* in the response to stress. Also, the presence 7 genes related to immune system processes (GO:0002376, they are *lola*, *pvf3*, *pvf2*, *lama*, *CG12780*, *ser* and *alpha-Man-1B*) and the fact *GATAe* has been itself previously associated with immune responses in the larval and adult MGs (Zhai et al., 2018, Senger et al., 2006). Given the vital roles of the MTs in immune response (Dow and Davies, 2006, McGettigan et al., 2005), and the implications of some of the detected targets in immune responses in other tissues (Tokusumi et al., 2018, Kleino et al., 2005), including *GATAe* (Senger et al., 2006) it would be compelling to investigate the roles of *GATAe* and its targets in the regulation of immune responses in the MTs.

There is a significant presence of genes involved in transport (GO:0006810, 19 hits, Figure 6.5.5B). More specific transport functions include transmembrane transport (7 hits), vesicle-mediated transport (6 hits), among others. According to FlyAtlas, 11 of the genes involved in transport are enriched in the MTs in normal conditions (Table 6.5-3), whereas 9 of them are not present (*Apoltp*, *best4*, *ctr1c*, *sur*, *task7*, *sytB*, *brp*, and *ninaC*), suggesting that *GATAe* could act both as an activator and as a repressor. Moreover, 9 of them are classified as transmembrane transporters (GO:0022857) according to their molecular function (*Best4*, *Ctr1C*, *CG6125*, *Glut1*, *Ir21a*, *Mdr65*, *Sur*, *Task7*, *Tim17b1*). This presence of transporters in the MTs correlates with the crucial transport functions performed by this organ (Beyenbach et al., 2010, O'Donnell et al., 1998, Cabrero et al., 2014). Also, it has been linked that the modulation of the expression of several MT transporters is associated with the response to stress (Stergiopoulos et al., 2009, Davies et al., 2014a). These data could potentially place *GATAe* as a regulator of transport genes in response to stress, which also

fits with the results shown in 4.3 that demonstrate lower tolerance to stress upon *GATAe* knockdown.

Gene	Adult Male		Adult Female	
	FPKM	Enrichment	FPKM	Enrichment
<i>CG6215</i>	228	17	419	115
<i>Scp2</i>	422	15	394	54
<i>Mdr6</i>	31	6.2	15	7.5
<i>GLUT1</i>	53	3.9	51	4.4
<i>Pld</i>	16	2.5	17	1.6
<i>Rlip</i>	8.1	1.6	11	1.7
<i>Vps16A</i>	9.6	1.5	9.9	1.2
<i>Zyx</i>	35	1.4	33	2.2
<i>CG5142</i>	2.7	1.4	6.6	3.3
<i>Tim17b1</i>	4.7	0.4	2.4	1.2
<i>Cadps</i>	4.8	0.3	2.5	0.8

Table 6.5-3. Genes involved in transport called as present in the MTs.

The table shows the genes involved with transport (according to *GOTermMapper*) detected in the ChIP-Seq analysis, and present in the adult MTs, scored by their enrichment. Adapted from FlyAtlas2. The abundance of each gene (FPKM), and enrichment, which measures the abundance of the gene relative to that in the whole fly, are indicated in the table.

In addition, both GO software detected that a significant part of the genes is involved in developmental processes (16 genes associated to “organ development” according to PANTHER, Figure 6.5.5A, and 24 to “anatomical development”, as reported by *GOTermMapper*, Figure 6.5.5). Again, this associates with and reinforces the roles in maintaining MT architecture shown in this thesis (see 4.2.2), and previous literature that associated *GATAe* with the development of the MG (Okumura et al., 2005, de Madrid and Casanova, 2018, Buchon et al., 2013). Also, the enriched GO terms of “cell communication” and “signal transduction” fit well in the model proposed here, in which *GATAe* is essential for MT growth and stem cell proliferation control (see 4.2.6).

Possibly, one of the most interesting genes detected in this category is *zyx*, an actin cytoskeleton component of the Hippo pathway, involved in tissue growth, which can regulate growth control in a Hippo dependent and independent manner (Gaspar et al., 2015, Rauskolb et al., 2011). Also, *zyx* functions are dependent on its interaction with cytoskeletal-binding protein Enabled (Gaspar et al., 2015), which, interestingly, is also expressed in the MTs (FlyAtlas2). Furthermore, it has been demonstrated that cell-to-cell communication is crucial for the maintenance of diverse stem cell populations including the

Drosophila ISCs and RNSCs (Singh et al., 2007, Ohlstein and Spradling, 2007). Importantly, many types of human cancers are induced by problems in processes of cell communication and signalling, and these processes are the focus for diverse therapies against this disease (Ghajar et al., 2013, Sinyuk et al., 2018, Oktay et al., 2015). Therefore, problems in cell communication could be one of the causes of the robust defects found upon *GATAe* knockdown in the PCs, and it would be of interest to study possible further interactions between *GATAe* and the Hippo pathway. Altogether, the analysis of the biological processes of this preliminary ChIP-Seq experiment matches closely with the model proposed in this thesis, in which *GATAe* is required to regulate cell-to-cell communication, signalling, and stress response processes in the PCs. Other important GO terms enriched in the detected peaks (according to *GOTermMapper*) were “cell proliferation” (5 hits, GO:0008283) and “cell cycle” (5 hits, GO:0007049). For example, these data suggest that *GATAe* can bind to the regions neighbouring *pvf2*, which is an essential factor in the age-induced proliferation of adult ISCs (Choi et al., 2008). As *pvf2* is not expressed in the adult MTs (FlyAtlas2), it could be possible that *GATAe* represses the expression of *pvf2* to control proliferation in the MTs. It is curious to note that, in human prostate cancer cell lines, both *GATA1* and *VEGFC* (the human homolog of *pvf2*) show an altered expression (Dozmorov et al., 2009) suggesting an interaction between these two gene families in human cancers. Another example is *xrp1*, which its possible role in the MTs is commented in 6.5.5.6.

Genes were also classified depending on their assigned molecular function/s. As well as shown with their biological processes, both GO tools classified the genes in similar groups according to their molecular functions. 9 genes are TFs (Figure 6.5.6A), and a similar number of them have DNA binding functions (11 genes, Figure 6.5.6B), suggesting that *GATAe* could be acting through the regulation of other TFs and DNA binding proteins in the PCs. Interestingly, a significant number of genes perform ion binding functions (5 genes involved in calcium ion binding Figure 6.5.6A, and 16 in ion binding, Figure 6.5.6B), and transport, which suggests a possible link between *GATAe* and the transport functions of multiple ions and compounds already reported in the MTs (Beyenbach et al., 2010, Wang et al., 2004). Other molecular functions associated with detected genes include growth factor activity and oxidoreductase activity.

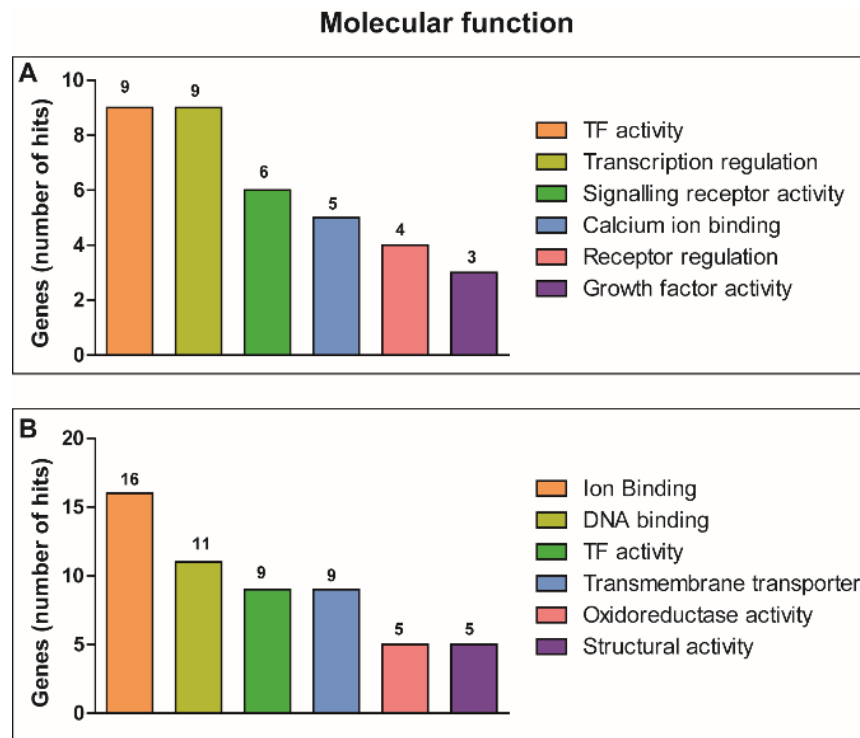


Figure 6.5.6. Significant ChIP-Seq detected genes classified according to their molecular function.

Graphs showing the distribution of genes classified according to their molecular function using either (A) PANTHER or (B) *GOTermMapper*. The number of genes that correspond to each category is indicated above every bar.

6.5.5.6 Top genes for further investigation

All genes associated with the peaks detected in this first ChIP-seq experiment are shown in Table 7.3-3. The broad classifications performed using PANTHER and *GOTermMapper* are useful to identify the core pathways that could be altered when *GATAe* expression is impaired. However, they do not explain, at this stage, which are the exact pathways in which *GATAe* is acting in the MT PCs for its maintenance and function. To do this, further investigation needs to be performed, and the results of this first sequencing can serve as an excellent start to address these questions, as described below.

It is important to consider that, given the number of dissections required to collect enough DNA for ChIP-seq, only one replicate was sequenced and analysed. Therefore, the conclusions extracted from the analysis of these data should be considered as preliminary, and as an initial step for future ChIP-seq experiments. However, this experiment still provides valuable data to identify novel functions of genes downstream of *GATAe*. For instance, one of the most reliable peaks obtained from this analysis (with an FDR of 2.52×10^{-30} , Table 7.3-3) detected in this experiment is *Xrp1*. It has been previously demonstrated

that *Xrp1* prevents genomic instability and cell proliferation in S2 cells (Akdemir et al., 2007). Given that this gene is also strongly enriched in the larval and adult MTs, it would be interesting to investigate if it also performs a function in the control of proliferation in the MTs, and it is indeed acting downstream of *GATAe*. Other possible targets identified in this analysis are well-studied genes, such as *serrate* (*Ser*), a well-known component of the Notch signalling pathway involved in numerous developmental processes (Pérez-Gómez et al., 2013, Thomas et al., 1991), or the TF *forkhead* (*fkh*). In the case of *fkh*, its functions in the embryonic development of the MTs are well known (Maruyama et al., 2011, Murakami et al., 1999, Gaul and Weigel, 1990). Also, *fkh*, is still highly enriched in the adult MTs (Table 6.5-4, (Wang et al., 2004)), and both genes perform similar roles for ISC maintenance (Lan et al., 2018). Furthermore, GATA factors have been previously related to ageing, and possibly interacting with *fkh* (Dobson et al., 2018, Alic et al., 2011), an event that has been already demonstrated in *C. elegans* (Tullet et al., 2008, Zhang et al., 2013). It would be compelling to investigate if *fkh* could perform novel roles in the adult renal system, aside from its developmental functions.

Furthermore, several unnamed genes were identified in the hit list of the ChIP-seq (a total of 49 different unannotated genes, Table 7.3-3). Interestingly, 9 of those genes exhibited specific enrichment in the larval and adult MTs, according to FlyAtlas2. Those genes are: *CG9932*, *CG10513*, *CG12780*, *CG13315*, *CG15093*, *CG17029*, *CG3332*, *CG5399* and *CG6125* (Table 6.5-4). Upon confirmation with additional ChIP-seq and qPCR experiments, it would be extremely valuable to identify how many of those genes are required for the maintenance or physiology of the MTs and study them in detail.

Gene	Adult Male		Adult Female		Larval		Molecular function
	FPKM	Enrichment	FPKM	Enrichment	FPKM	Enrichment	
<i>Fkh</i>	38	7.5	44	22	18	7.5	TF
<i>Hebe</i>	52	2.7	65	6.2	30	1.8	Unknown
<i>Xrp1</i>	148	2.6	147	7.3	95	3.0	P-element binding
<i>CG9932</i>	37	1.4	33	3.3	8.5	1.2	TF
<i>CG10513</i>	1456	32	926	91	2949	32	Unknown
<i>CG12780</i>	118	1.1	197	7.3	0.2	0.0	Carbohydrate binding
<i>CG13315</i>	4051	3.6	2006	5.9	491	1.5	Unknown
<i>CG15093</i>	432	7.4	339	8.7	154	5.1	Oxidoreductase activity
<i>CG17029</i>	569	14	620	35	543	13	Inositol monophosphate 1-phosphatase activity
<i>CG3332</i>	12	2.3	21	5.2	13	1.1	Unknown
<i>CG5399</i>	263	2.4	387	4.3	86	0.3	Unknown
<i>CG6125</i>	228	17	419	115	83	10	SLC26A/SulP transporter

Table 6.5-4. Table of the most tubule-enriched genes from the ChIP-seq analysis.

The table includes three enriched known genes (*fkh*, *hebe* and *Xrp1*), and all the unnamed CG genes reported in Table 7.3-3 that are specifically enriched in either larval or adult MTs (both males and females). The abundance of each gene (FPKM), and enrichment, which measures the abundance of the gene relative to that in the whole fly, are indicated in the table. Adapted from FlyAtlas 2. Molecular functions according to NCBI, Flybase, and previous literature are shown in the last column.

Curiously, none of the genes examined in the qPCR experiments of this thesis (Figure 4.2.16) was detected in this ChIP analysis, suggesting that the variation in their expression could potentially be an indirect consequence of *GATAe* loss in the PCs, rather than being directly being regulated by *GATAe*.

Altogether, this first ChIP-seq experiment provides valuable information on which genes could be acting downstream of *GATAe* in the MTs, and in which processes are they involved in order to maintain the integrity of the adult MTs. Therefore, the data presented here, although preliminary, could serve as a potentially good base to study novel genetic interactions of *GATAe* in the MTs and decipher the mechanisms of *GATAe* function in the maintenance of the homeostasis of the adult fly renal system.

6.6 Discussion and troubleshooting

In this chapter, different molecular tools have been elaborated and tested for functional characterisation of *GATAe*. A polyclonal antibody against *GATAe*

protein was produced, and an efficient *GATAe* gain-of-function line, containing an HA tag was generated and used for further qPCR and ChIP experiments. Unfortunately, the generated antibody does not specifically recognise *GATAe* protein, and as such is unsuitable for subsequent use in WB or ChIP experiments. However, the *GATAe*-HA overexpression fly line expressed the HA-tagged form of *GATAe* in the nuclei of the MTs and was successfully used for ChIP-seq.

6.6.1 Further experiments to assess *GATAe* molecular function

The morphological and functional defects induced by silencing *GATAe* in all MT cell types have been characterised in previous chapters. However, the potential effects of *GATAe* overexpression in the MTs merits further attention.

Overexpression of *GATAe* in PCs did not result in any particular morphological defects (although driving this construct using *ctB>Gal4* induced embryonic lethality), but the accumulation of potential structures resembling kidney stones was detected mainly in the initial segments of the *CapaR>UAS-GATAe*-HA MTs (see Figure 6.4.4). Overexpression of *GATAe* in SCs or RNSCs has not been examined in detail regarding fluid secretion or cell survival, but ICC experiments in *CIC-a>UAS-GATAe*-HA reveal normal morphology of the MTs (Figure 6.4.5).

Furthermore, rescue experiments using the *UAS-GATAe*-HA line would be valuable to show that *GATAe* mutant lethality can be reversed by reintroducing a copy of the *GATAe* gene. However, all *UAS-GATAe*-HA lines generated are inserted in the third chromosome, and since *GATAe* is also localised in this chromosome, rescue experiments using any of the previously generated mutant lines *GATAe*⁻ (Okumura et al., 2016, de Madrid and Casanova, 2018), are difficult to perform. Therefore, obtaining a homozygous mutant for *GATAe* containing at least a copy of *UAS-GATAe*-HA could be generated through genetic recombination, and it would be of extreme interest to combine both constructs perform future rescue experiments.

In addition, as explained in 6.4.2, rescue experiments were performed using the RNAi line to silence the *GATAe* expression, and the overexpressing construct *UAS-GATAe*-HA. However, the MTs phenotype obtained in *GATAe* knockdown flies could not be rescued using the *GATAe* gain-of-function line. Also, as commented previously, one possible cause is that RNAi silencing effects are

epistatic over *GATAe* overexpression. Thus, although *GATAe* expression would be induced at very high levels in UAS-*GATAe*-HA MTs, at the same time, RNAi would bind to all *GATAe* mRNA, resulting in low levels of *GATAe* protein. In order to obtain flies that overexpress the *GATAe*-HA construct in the PCs, a recombinant fly line which contains 2 copies of both *CapaR*-Gal4 and UAS-*GATAe*-HA constructs was recently generated in this project (*CapaR*;UAS-*GATAe*-HA, Figure 6.6.1, Dr Selim Terhzaz). As both *CapaR* and UAS-*GATAe*-HA constructs are already in the same chromosome, this doubly-homozygous line will facilitate the collection of considerable number of flies of the desired genotype for further ChIP experiments.

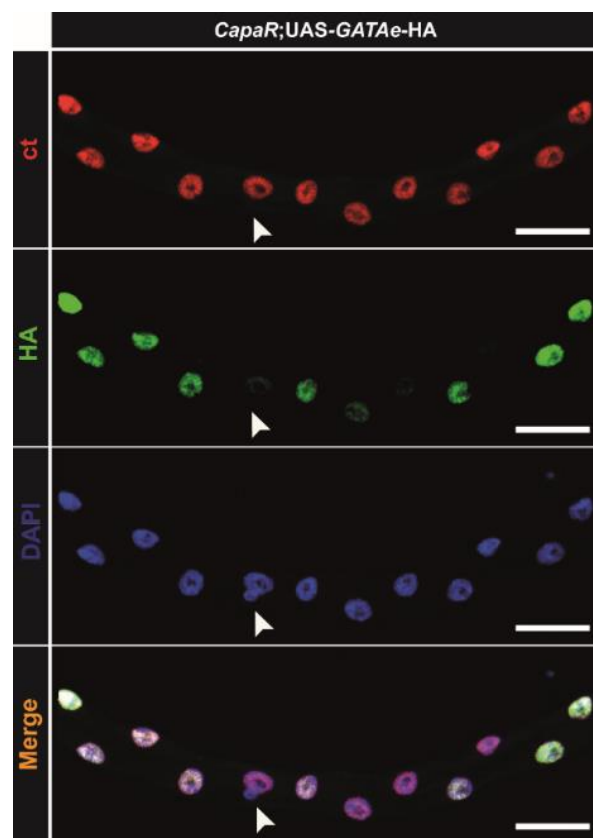


Figure 6.6.1. Immunocytochemistry of recombinant *GATAe* to PCs

Recombinant *CapaR*;UAS-*GATAe*-HA adult MTs expressing *GATAe* specifically in PCs were immunostained with ct (red), HA (green) and DAPI to confirm HA expression in the nucleus of PCs. Note the absence of HA staining in SCs (white arrow). Scale bars are 50µm.

Furthermore, the ChIP-seq results obtained in this project using the *GATAe*-HA line are promising, and further reinforce the roles of *GATAe* in the PCs in cell communication, signalling and developmental processes. However, and as mentioned in 6.5.5.3, it is crucial to consider that this data is extracted from only one replicate, and it should be analysed with caution. Additional replicates need to be performed to confirm these results, and further ChIP-qPCR

experiments should be done to confirm the binding sites of *GATAe* in the genome. In parallel, measuring mRNA expression of the top candidates from this ChIP-seq experiment could be performed in adult *CapaR:UAS-GATAe-HA* MTs to further confirm these hits.

Also, an alternative to ChIP-seq could be the generation of transgenic line containing a Gal4-inducible form of *GATAe* fused to a bacterial DNA methylase (*UAS-GATAe-Dam*). Using this technique, called Targeted DamID (TaDa), genome-wide profiling of *GATAe* could be investigated employing far fewer number of cells, compared to other methods (including ChIP-seq) (Marshall et al., 2016, Southall et al., 2013). A crucial advantage of the TaDa technique is that it does not require a specific antibody that binds to the protein of interest, avoiding the potentially problematic step of generating a specific antibody for *GATAe*. In addition, unlike ChIP-seq, TaDa does not require fixation of the sample, facilitating the extraction of DNA in unfixed or even live cells (Aughey and Southall, 2016).

Precisely, the *UAS-GATAe-Dam* construct could be expressed in each of the three cell-types of the MTs (PCs, SCs and RNSCs), by using cell-specific Gal4 lines. In fact, the same Gal4 lines that have been used in this study could be employed (*CapaR-Gal4* for PCs, *clc-a-Gal4* for SCs, and *esg-Gal4* for RNSCs), combined with a *Gal80^{ts}* construct. This would allow the examination of genome-wide occupancy of *GATAe* in the adult stage (or in any other developmental stage) independently in every single MT cell-type. Once the direct downstream genes of *GATAe* are identified, their expression could be manipulated (by RNAi, or overexpression), in an attempt to recapitulate, or even rescue, the phenotype observed upon *GATAe* RNAi. This would potentially provide an exact picture of the function of *GATAe* in each MT cell-type.

Altogether, this first replicate serves as a solid base to perform these additional experiments (e.g. it provided possibly new genes to perform ChIP-qPCR, and it has been crucial to confirm that the ChIP protocol used here is valid).

In summary, a gain-of-function line for *GATAe* has been successfully generated and tested in this study. Preliminary results of ChIP-Seq analysis using this line driven in the PCs suggests that *GATAe* is acting upstream of several pathways

(developmental, cell communication and signalling pathways) to ensure the morphology and function of the MTs. Also, potential novel genes that could have a role in the function of the adult MTs have been identified in this ChIP-Seq experiment. Thus, these results are an excellent starting point for further experimentation that is required to decipher the molecular mechanism of the function of *GATAe* in the MTs.

Chapter 7 - Summary and future work

7.1 Summary

The primary goals of this thesis (shown in 1.5) were to design a complete genetic screen of TFs that may have a novel function in the MTs of *Drosophila*. After this screen, the TF *GATAe* was selected for further study, with the objective to decipher its novel roles in the MTs, by a diverse range of techniques, including the generation of new antibodies and tagged lines. A secondary and independent goal of this project, in collaboration with the RENALTRACT consortium, was to design a protocol to obtain high-quality 3D images of whole-fly preparations. All the aims of this PhD project, outlined in section 1.5, have been extensively covered and investigated, and the experiments and results of this thesis are summarised in here.

7.1.1 Genetic screen

By performing this screen, it was possible to reduce the original list of 30 TF enriched in the MTs to a list of 10, by searching the literature, documenting their patterns of expression and their human homologues. Of these 10 TFs, *GATAe* was selected to be investigated in detail.

However, the screen also resulted in three other TFs - *ptx1*, *HNF4* and *hth*- as strong candidates as novel regulators of development or physiology of the MTs. This is due to their enrichment during all the stages of development, their human counterparts involved in kidney development, or the lethality phenotypes observed upon downregulation of those genes in the MTs.

In addition, apart from the TFs mentioned above, the ECM component *Muc11A* is another potential candidate to function in the development and morphology of MTs, as demonstrated by ICC experiments (see 3.3.1.1).

7.1.2 *GATAe* involvement in the MTs

The function of *GATAe* in MTs has been characterised in detail in this PhD project. Although *GATAe* is expressed in the MTs at all developmental stages, this is the first report to date that shows the functional involvement of *GATAe* in the MTs. *GATAe* is required in all three types of renal cells but is playing very different roles in each of them.

7.1.2.1 *GATAe* in the PCs

The results showed in Chapter 3 demonstrate that *GATAe* is specifically expressed and necessary in the PCs to control tissue growth and cell proliferation. Downregulation of *GATAe* in the PCs (using diverse and independent RNAi and Gal4 lines) induce severe defects in the morphology and function of the adult MTs (see 4.2).

However, this is not a consequence of MT developmental effects, since Gal80^{TS} experiments showed that only when *GATAe* expression is silenced from L3 onwards, that MTs start displaying morphological abnormalities, such as tumourous growth (see 4.2.4) or uncontrolled proliferation of potential RNSCs (see 4.2.6). qPCR experiments also demonstrated that *GATAe* downregulation in the PCs alters the expression of specific apoptosis and cell proliferation-related genes, as well as oncogenes and tumour suppressor genes, suggesting that *GATAe* can act as a master regulator upstream of all these genes (see 4.2.7).

7.1.2.2 *GATAe* in SCs

GATAe is also required for the survival of the SCs throughout development. Downregulation of *GATAe* in the SCs results in a significant and robust reduction in SC population, and the few SCs observed in these MTs are restricted to the initial segment, being absent in the main segment. In addition, these MTs showed impaired fluid secretion upon stimulation with kinin, a neuropeptide that stimulates specifically secretion through SCs (see 5.2). Thus, these results indicate that *GATAe* is required for the survival of SCs through metamorphosis.

7.1.2.3 *GATAe* in RNSCs

Apart from its functions in PCs and SCs, *GATAe* is also required for the early migration of the RNSCs during metamorphosis. Silencing *GATAe* in the RNSCs exclusively (by using the *esg>GFP^{ts}* line) is sufficient to deplete the whole RNSC population in the adult ureters and lower tubules (see 5.3). However, *GATAe* seems to be acting in the early migration of the RNSCs, as when *GATAe* was silenced from 24h APF, the RNSCs were correctly located in the adult MTs (also shown in 5.3). Furthermore, *GATAe* silencing in adult RNSCs did not induce any

loss of this cell population, suggesting that *GATAe* is required for RNSC migration rather than their adult maintenance.

Altogether, these results indicate that *GATAe* is required in all three MT cell-types (PCs, SCs and RNSCs), being involved in different processes for each of these cell types. A proposed model for the novel roles of *GATAe* in the MTs is shown in 5.5.

7.1.3 Molecular and biochemical tools for *GATAe* functional characterisation

7.1.3.1 Antibody and gain-of-function line generation

To further investigate the mechanism in which *GATAe* modulates MT morphology and physiology, an antibody against the *GATAe* protein was generated. Theoretically, this antibody would have been used to confirm the pattern of expression of *GATAe*, and potentially to perform ChIP-seq experiments to identify novel downstream targets of *GATAe*. However, the experiments performed using this antibody demonstrated that it is not suitable for these purposes, for two main reasons: First, ICC experiments using different developmental stages of MTs were not able to detect *GATAe* protein in the cell nuclei using this antibody. Second, Western-blot analysis confirmed that the antibody generated was not specific for the *GATAe* protein (see 6.3).

In addition, a gain-of-function *GATAe* line tagged with HA was generated in this study. It was extremely useful to investigate the overexpression phenotypes and to confirm downstream targets of *GATAe* (e.g., *intB* or *inx7*) *GATAe* overexpression in the PCs of adult MTs induces the accumulation of aggregates in the tubule lumen, especially in the initial segment (see 6.4).

7.1.3.2 ChIP-Sequencing

In order to identify direct downstream targets of *GATAe* in the PCs, adult *CapaR>UAS-GATAe-HA* MTs were analysed by ChIP-seq. The preliminary results using this technique showed that the *GATAe* protein binds specifically near to at least 108 different genes, in the cellular context of the adult (5-7 d.o.) PCs (Appendix 10). A significant number of these genes were associated with

metabolic processes, such as nitrogen metabolism. Also, other biological processes overrepresented in these genes include development, cell differentiation and cell communication processes, which could explain the phenotypes observed upon *GATAe* downregulation in the PCs.

In addition, almost half of the genes detected (49 out of 110) are still unnamed (CG number identification), indicating that most of them could have a novel role in the MTs. At least 9 of them are significantly enriched in the adult MTs, according to FlyAtlas 2. Therefore, these genes deserve further attention and would be interesting to explore their potential roles downstream of *GATAe* (see 6.5.5.6).

7.1.4 Observation of 3D structures in whole-fly preparations

This part of the project focused on the elaboration of a protocol to observe intact MTs in undissected flies. To do this, different microscopy techniques were employed. This part of the project disclosures a new way to observe whole-fly structures and organs without the requirement of dissections, and with minimal alteration of organ position. Many aspects of this protocol, however, need to be still improved, such as a more effective way to permeabilise the adult cuticle without manual removing. Other parameters such as increasing the fixation time of concentration, or CUBIC treatment duration could as well influence in the quality of the images.

7.2 Future work

The results obtained in this PhD thesis have been generally successful and uncovered new TFs and other genes that could perform novel roles in the MTs. Nevertheless, and as mentioned in previous sections, there are several questions that were raised during the course of this thesis which could be potentially interesting to answer in future investigations. This section will try to summarise the remaining open questions and which directions could be explored to answer these.

7.2.1 Study of other genes detected in the screening

As mentioned in 3.3.1 several other genes aside from *GATAe* were identified via the genetic screening. Several lines of research could emerge from these results, investigating any of the other exciting genes identified during this selection process. Perhaps the most interesting gene to be studied could be *Muc11A*. Low levels of *Muc11A* in the PCs induced severe morphological defects (see Figure 3.2.1B). In addition, the emerging importance of the mucin family in the development of diverse tissues of both vertebrates and invertebrates (Syed et al., 2008, Dhanisha et al., 2018), makes this gene (and possibly other genes of the mucin family) a motivating target to study in the MTs.

Another potential gene which may modulate novel functions in the MTs is *HNF4*. It has been previously reported that *HNF4* gene is enriched in the MTs, (Palanker et al., 2009) in accordance to FlyAtlas data, and its direct human homolog, HNF4, is also expressed in the kidney in an isoform-specific manner (Duncan et al., 1994, Dean et al., 2010). These data suggest that *HNF4* could potentially have a function in the developing or adult MTs, and these functions may be conserved along evolution. Therefore, further research on *Drosophila HNF4* could shed light on a renal role.

7.2.2 Further research on *GATAe*

Regarding *GATAe*, there are several directions to where subsequent research could be pointing to decipher its functions in the renal system and its possible homologies in vertebrates.

7.2.2.1 ChIP-Sequencing

An exciting achievement of this research project has been the generation of a functional overexpression line for *GATAe*, which contains an HA tag. Therefore, the immediate future work that should be done is to continue the ChIP-Seq experiments. The ChIP-seq data presented in 6.5, although promising, is preliminary, as only one replicate was sequenced. Therefore, a good starting point for future experiments would be to use the recombinant-generated line *CapaR>UAS-GATAe-HA* to perform further replicates of the ChIP-seq experiments and identify the direct downstream targets of *GATAe* in the MTs. Also, ChIP-

qPCR of the top hits should be performed to confirm the gene hits found. This would then further elucidate those specific pathways modulated by *GATAe* to maintain MT architecture. Also, alternative approaches to identify downstream targets of *GATAe* could be performed to confirm the ChIP-seq data. RNA-Seq experiments would provide useful insights in which genes show altered expression upon downregulation or gain-of-function of *GATAe*. However, this technique would not provide information about which of them are direct targets. Alternatively, the targeted DamID (TaDa) technique has been proven to be very useful to identify chromatin binding sites (Marshall et al., 2016). This technique allows the detection DNA-binding sites of proteins in a rapid way without fixation and IP steps and using very low initial quantity of tissue or cells (Marshall et al., 2016). Thus, it could be a valuable alternative of ChIP-Seq or an additional way to confirm the hits found using this technique.

7.2.2.2 *GATAe* in RNSCs

Possible novel roles of *GATAe* in the RNSCs have been demonstrated in this thesis (see 5.5). The results showed here suggested that *GATAe* is not required for RNSC survival, as happens with other factors, such as *yorkie* or *svb* (Bohère et al., 2018). However, low levels of *GATAe* in RNSCs before their migration to the MTs caused a total absence of RNSCs in the adult ureters, suggesting a role for *GATAe* in RNSC integration to the MTs. However, many additional questions arise from these results. Is *GATAe* involved in the active division rates of the RNSCs? It has been shown that *GATAe* controls the proliferation of ISCs (Okumura et al., 2016). It could be possible that *GATAe* performs similar functions in the RNSCs, but these roles could be masked due to the extremely low mitotic rates of RNSCs (Li et al., 2015, Zeng et al., 2010). Also, the phenotypes observed here upon *GATAe* downregulation in RNSCs are strikingly similar to the ones observed when *Rac1* function is impaired (Takashima et al., 2013), suggesting a possible interaction between these two genes. Thus, it would be worth investigating if there is a genetic or physical interaction between these two proteins to pinpoint the function of *GATAe* in RNSCs.

7.2.2.3 Conservation of GATA factor functions

Finally, it would be compelling to investigate possible homologies of GATA factors in the development and adult maintenance of the vertebrate kidney. Significant evolutionary similarities have been presented from the results of this thesis (e.g. the involvement of human GATA factors in tumorigenesis, or even the interactions between the GATA and bcl protein families, see 4.2.7). Therefore, one of the best directions that subsequent research could take is to investigate, for example, if introducing a copy of the *Drosophila* GATAe in vertebrate kidney models (e.g. *Xenopus* embryos or mouse kidney) can rescue some phenotypes observed induced by loss of vertebrate GATA factors.

7.3 Final conclusion

In conclusion, the data presented in this thesis can serve as a strong base to investigate novel functions of genes involved in the development and homeostasis of the *Drosophila* renal tubules. Strikingly, the results obtained with the study of GATAe in the MTs has uncovered new functions for GATAe in all three cell types of the MTs, thus reinforcing the functional versatility of this TF depending on the specific cellular context. This, together with further research in *Drosophila* and additional systems (perhaps mammalian or organoid models) would provide new and vital insights into the mechanisms of development and adult maintenance of the kidney.

Appendices

Appendix 1 *Drosophila* media composition

Compound	Quantity
Soya Flour	5.25g/l
Maize Meal	15g/l
Yeast	35g/l
Wheat germ	10g/l
Glucose	30g/l
Treacle	30g/l
Sucrose	15g/l
Agar	10g/l
Propionic acid	5ml/l
Methylparaben	10ml/l
Water	Up to required

Table 7.3-1. Recipe for the *Drosophila* Standard medium (Fly food).

Appendix 2 Grape juice agar medium preparation

Compound	Quantity
Agar	20g/l
Glucose	52g/l
Sucrose	26g/l
Dried yeast	7g/l
Red grape/orange juice	88ml/l
Nipagin	7ml/l
Water	Up to required

Protocol for embryo collection medium preparation:

1. Dissolve agar in ~800ml of water.
2. Add the dried yeast and bring to boil.
3. Once the solution is boiling, add the glucose and sucrose.
4. Boil for 10min
5. Add the red grape/orange juice, and cool down.
6. Once the solution is cooled down, add the Nipagin
7. Add water up to 1L and pour in the plates.

Any remaining medium that has not been poured into the plates can be stored at 4°C in a 500ml sterile glass bottle for up to 6 weeks.

Appendix 3 E. coli growth media recipes

L-broth	per litre of water
	10g Bacto-tryptone
	5g yeast extract
	10g NaCl
L-agar	per litre of water
	10g Bacto-tryptone
	5g yeast extract
	10g NaCl
SOC medium	15g Bacto-agar
	2% tryptone
	2.5mM KCl
	10mM MgCl ₂
	10mM MgSO ₄
	20mM glucose

Appendix 4 List of primers used

All primer pairs used have been designed using MacVector™ 14.5.3 Software (MacVector Inc.) and summarised in Table 7.3-2.

Primer pairs	Sequence 5'-3'	Purpose
rpl32F	TGACCATCCGCCCAGCATAC	Housekeeping gene for PCR and qPCR. Test the expression of <i>rpl32</i> .
Rpl32R	ATCTCGCCGAGTAAACG	Housekeeping gene for PCR and qPCR. Test the expression of <i>rpl32</i> .
DIAP1F	CGTGGTGCATAAGAGGTGA	Used in qPCR to test the expression of <i>DIAP1</i> .
DIAP1R	TTGAATAGCTGGGTGCGTT	Housekeeping gene for PCR and qPCR. Test the expression of <i>rpl32</i> .
Ras85DF	GCAAGAGAGGTGGCCAAACA	Used in qPCR to test the expression of <i>Ras85D</i> .
Ras85DR	TCGGCTTGTTTCATTTGCGG	Used in qPCR to test the expression of <i>Ras85D</i> .
wtsF	AGCCGACAATAACTGGGTGG	Used in qPCR to test the expression of <i>Warts</i> .
wtsR	CGAGTGATTGCCGTTCTCT	Used in qPCR to test the expression of <i>Warts</i> .
debclF	TTTTTCGCTCCAGCATCACC	Used in qPCR to test the expression of <i>debcl</i> .
debclR	CGTCAATCCCAAGAACG	Used in qPCR to test the expression of <i>debcl</i> .
Ahcyl1F	GGCGAGACGGAAGAGGACT	Used in qPCR to test the expression of <i>Ahcyl1</i> .
Ahcyl1R	AGAGAGCTGATAGAGACGGTG	Used in qPCR to test the expression of <i>Ahcyl1</i> .
BuffyF	GCCACACTACATTCCGCATCAC	Used in qPCR to test the expression of <i>buffy</i> .
BuffyR	ATTCATCGCCAGCACTTC	Used in qPCR to test the expression of <i>buffy</i> .
Inx7F	GGTGTGTGTCTCCGGTCAT	Used in qPCR to test the expression of <i>GATAe</i> .
Inx7R	CACAGGGCATGGGAATGTA	Used in qPCR to test the expression of <i>GATAe</i> .
IntBf	ACGCACTCAGTTTGCCGATA	Used in qPCR to test the expression of <i>IntB</i> .
IntBR	TGTTTCCTGACAGGGACACG	Used in qPCR to test the expression of <i>IntB</i> .
ScribF	AGACGCAAACGACCTCCACG	Used in qPCR to test the expression of <i>scribbler</i> .
ScribR	CCTGGGTGGCTTTGCCAGAA	Used in qPCR to test the expression of <i>scribbler</i> .
GATAe1F	AAGAAACGCAAAGAAGCGGC	Used in qPCR to test the expression of <i>GATAe</i> .
GATAe1R	TTAATTCAAGTGTGCCGGCTG	Used in qPCR to test the expression of <i>GATAe</i> .
GATAe2F	GTTTCAGCCAGTACCCACCAT	Used in qPCR to test the expression of <i>GATAe</i> , and for sequencing UAS- <i>GATAe</i> -HA.
GATAe2R	CTGGGACATTGGTCTTGGC	Used in qPCR to test the expression of <i>GATAe</i> , and for sequencing UAS- <i>GATAe</i> -HA.

<i>GATAe3F</i>	ATGGTCTGCAAACTATCTCACCG	Used in qPCR to test the expression of <i>GATAe</i> , and for sequencing UAS- <i>GATAe</i> -HA
<i>GATAe3R</i>	TTCGCTGACGCCCCTTGGCCCG TCT	Used in qPCR to test the expression of <i>GATAe</i> , and for sequencing UAS- <i>GATAe</i> -HA

Table 7.3-2. Primer sequences and their purpose

Appendix 5 5X TBE buffer recipe

Compound	Quantity
Tris base	54g/l
Boric acid	27.5g/l
EDTA pH8	20ml/l

Make up to 1L with ddH₂O and then dilute 1/10 for 0.5X working solution.

Appendix 6 10X PBS and 10X TBS recipes

For 10X PBS (Phosphate-Buffered Saline)

Compound	Quantity
NaCl	80g/l
KCl	2g/l
Na ₂ HPO ₄	14.4g/l
KH ₂ PO ₄	2.4g/l

Dissolve all components in 800ml ddH₂O and adjust pH to 7.4. Then make up to 1L. Dilute 1/10 for a 1X working solution.

For 10X TBS (Tris-Buffered Saline)

Compound	Quantity
Tris base	24g/l
NaCl	88g/l

Dissolve all components in 800ml ddH₂O and adjust pH to 7.4. Then make up to 1L. Dilute 1/10 for a 1X working solution.

Appendix 7 Solutions 1 and 2 for ECL development

Chemicals needed:

Luminol: a 250mM stock is prepared in DMSO and kept in the dark at -20°C in 100µl aliquots.

P-coumaric acid: a 0-mM stock is prepared in DMFO and kept in the dark at -20°C in 50µl aliquots

The composition of solutions 1 and 2:

<i>Solution 1</i>	<i>Quantity</i>
<i>250mM Luminol</i>	100µl
<i>90mM P-coumaric acid</i>	44µl
<i>1M Tris-HCl pH 8.5</i>	1ml
<i>ddH₂O</i>	Up to 10ml
<i>Solution 2</i>	
<i>H₂O₂</i>	6.4µl
<i>1M Tris-HCl pH 8.5</i>	1ml
<i>ddH₂O</i>	Up to 10ml

Appendix 8 CUBIC solution preparation

Compound	Quantity
Urea powder	25g
Quadrol (Sigma)	24.96g
ddH ₂ O	28.8g
Triton X-100	15g

Protocol for CUBIC solution preparation:

For 100g final weight, measure all components by weight, not by volume, due to the high density of Quadrol and Triton X-100.

1. Weight Quadrol
2. Add the urea powder and ddH₂O
3. Warm and stir for 10 min
4. Cool down the solution
5. Add the Triton X-100
6. Store at RT for several weeks

Appendix 9 *Drosophila* saline recipe

Recipe adapted from (Dow et al., 1994).

Dissolve in ~800ml of water

Compound	mM	g/l
NaCl	117.5	6.86
KCl	20	1.49
CaCl ₂ *H ₂ O	2	0.29
MgCl ₂ *6H ₂ O	2	0.41
Glucose	20	3.96
HEPES	8.6	2.05

Dissolve separately in 100ml of water each, and then add to the rest of the mixture

NaHCO ₃	10.24	0.86
NaH ₂ PO ₄	4.5	0.70

Adjust pH to 6.8 and add water up to 1l, filter sterilise and store at 4 °C

Appendix 10 ChIP-Sequencing gene list

Gene name/Symbol	Fold Change	P-Value	FDR-Threshold
Xrp1	2.61421	2.24E-33	2.53E-30
CG17029	2.482363	4.25E-27	3.43E-24
CG10513	2.256099	2.68E-20	9.48E-18
CG41520	2.32489	3.6E-17	9.68E-15
hebe	2.045903	9.38E-16	1.83E-13
CG17475	1.646703	3.23E-11	3.44E-09
Ser	1.604976	1.18E-09	1.09E-07
lola	1.581707	2.73E-09	2.31E-07
CG6125	1.618846	4.43E-09	3.58E-07
Adgf-E	1.554887	5.31E-09	4.17E-07
CG5399	1.534427	6.92E-09	5.15E-07
CG17514	1.538586	1.68E-08	1.19E-06
Pms2	1.5302	3.36E-08	2.26E-06
Sytbeta	1.416815	4.54E-08	3.02E-06
CG18577	1.48793	9.11E-08	5.48E-06
CG2111	1.555304	1.87E-07	1.03E-05
CG18666	1.503826	2.03E-07	1.09E-05
Tie	1.471477	2.64E-07	1.35E-05
CG12780	1.481234	2.66E-07	1.34E-05
retn	1.633881	3.99E-07	1.93E-05
CG17687	1.529261	5.14E-07	2.44E-05
mus205	1.519235	6.24E-07	2.87E-05
CG10559	1.513551	1.13E-06	4.8E-05
l(2)09851	1.499667	1.16E-06	4.88E-05
CG5823	1.589722	1.41E-06	5.62E-05
Cadps	1.342624	1.41E-06	5.59E-05
CG13315	1.473426	1.92E-06	7.45E-05
Ggamma30A	1.429009	2E-06	7.7E-05
CG15093	1.442194	2.31E-06	8.61E-05
Fmo-2	1.466874	3.42E-06	0.000124
CG44227	1.412284	3.52E-06	0.000127
Rlip	1.406559	3.98E-06	0.00014
fkf	1.400999	4.48E-06	0.000154
CG34265	1.453165	5.51E-06	0.000181
CG7695	1.399037	5.69E-06	0.000186
CG14853	1.445271	6.38E-06	0.000204
Best4	1.397016	7.22E-06	0.000224
CG17324	1.377853	7.38E-06	0.000227
lama	1.377853	7.38E-06	0.000227
ect	1.388825	8.56E-06	0.000257
CG12883	1.43971	8.76E-06	0.000262
brp	1.386174	9.05E-06	0.000269

CG31122	1.375942	1.12E-05	0.000323
TrissinR	1.373474	1.18E-05	0.000335
Zyx	1.381358	1.21E-05	0.000342
vg	1.415692	1.67E-05	0.000453
CG33143	1.368842	1.88E-05	0.000503
CG32816	1.489107	2.33E-05	0.000605
Ir21a	1.363936	2.5E-05	0.000637
Pld	1.347474	3.48E-05	0.000859
CG7213	1.345252	3.64E-05	0.000886
CG13693	1.351758	3.82E-05	0.000926
CG9932	1.347126	4.18E-05	0.001001
Pvf2	1.361313	4.56E-05	0.001066
CG14020	1.378384	4.86E-05	0.001125
ND-B14.5AL	1.388292	4.95E-05	0.001138
Vps16A	1.342171	5.49E-05	0.001242
Nse1	1.333365	6.52E-05	0.001428
CG3332	1.341697	6.61E-05	0.001439
CG31176	1.341697	6.61E-05	0.001439
Gr93b	1.334916	7.51E-05	0.001576
lgs	1.334916	7.51E-05	0.001576
euc	1.372909	7.81E-05	0.001623
Glut1	1.350616	7.98E-05	0.001654
Cpr47Ef	1.369866	8.22E-05	0.001678
Menl-1	1.32103	8.29E-05	0.001687
Cpr49Ad	1.332045	9.42E-05	0.001869
CG13438	1.300258	0.000107	0.002053
SdhB	1.363936	0.00011	0.002078
Syn2	1.355192	0.000127	0.00235
CG7101	1.313249	0.000134	0.002415
CG5142	1.320406	0.000138	0.002478
ninaC	1.3579	0.000145	0.002579
CG4631	1.354962	0.000153	0.002681
CG2217	1.351758	0.000193	0.003245
tnc	1.291476	0.000203	0.003355
MsR1	1.284927	0.00023	0.003676
CG1402	1.293932	0.000265	0.004121
alpha-Man-Ib	1.406559	0.000276	0.004283
Tim17b1	1.29809	0.000287	0.004378
sns	1.269871	0.000308	0.004582
slp2	1.331601	0.000317	0.004684
CG4669	1.283576	0.00032	0.004675
CG13891	1.290704	0.000328	0.004765
tna	1.283704	0.000372	0.005216
CG32052	1.321313	0.000374	0.005228
CG14546	1.272452	0.000393	0.005434
CG8929	1.273864	0.000444	0.005936
CG16779	1.257259	0.000453	0.006031

Task7	1.380369	0.000475	0.006262
Apoltp	1.275368	0.000502	0.006519
CG12617	1.360188	0.000513	0.00662
unc79	1.27374	0.000516	0.006633
Mdr65	1.308037	0.000544	0.00696
CG9624	1.303438	0.000585	0.007315
CG42355	1.313964	0.000586	0.007312
CG31176	1.257629	0.000597	0.007378
Pvf3	1.298986	0.000627	0.007659
Sur	1.268726	0.000653	0.007908
Act79B	1.259977	0.000659	0.007899
Ctr1C	1.294673	0.000671	0.007991
hng3	1.26714	0.000671	0.007975
Pif1A	1.264029	0.000709	0.008366
Sf3b2	1.225612	0.000737	0.008632
Scp2	1.26846	0.00076	0.008821
h-cup	1.290868	0.000834	0.009469
caup	1.348436	0.000866	0.009703
CG9040	1.348436	0.000866	0.009703
CG13643	1.26054	0.000869	0.009692
lds	1.26818	0.000884	0.00982

Table 7.3-3. Complete gene list detected by the ChIP-Seq.

The complete list of genes detected in the first replicate of the ChIP-seq experiment using *CapaR>UAS-GATAe*-HA adult MTs, immunoprecipitated using an antibody against the HA tag. Only genes detected using an FDR<0.01 (p-value adjusted by multiple testing employing the Benjamini-Hochberg method) are shown in this list (see methods and 6.5.5.3 for more information). Genes are ranked by their Fold change (comparing sample vs control). Genes highlighted in yellow are the ones that were found to be unannotated according to *GOTermMapper*.

Appendix 11 Observation of whole-fly Malpighian tubules using tomography projection microscopy

7.4 Introduction

The research presented here was performed in the laboratory of Professor Seppo Vainio, in BioCenter Oulu (Finland), as part of a secondment of the RENALTRACT network. This internship aimed to obtain whole-fly 3D images of adult and larval MTs through the cuticle without the need of dissection. The primary purpose during this visit was to: (i) design a protocol to fix and clear whole-fly larval and adult *Drosophila* samples. (ii) use OPT and two-photon microscopy technologies to obtain 3D pictures of these cleared samples, by detecting endogenous GFP expression.

It has been demonstrated that the migration of the MTs to their final positions in a highly stereotypical manner is crucial for the survival of the fly through development (Weavers and Skaer, 2013, Denholm, 2013). Whereas the migration of the developing MTs within the embryo can be easily observed using confocal microscopy, their localisation in later stages (L3, pupal, or adult stages) requires more sophisticated imaging systems, due to the increased size of the samples. Therefore, the experiments performed during this secondment could represent a good advance into the design of new protocols to observe whole-fly adult organs. It is particularly important to study these processes in the MTs since they change their shape and length dramatically during metamorphosis dramatically (see 1.2.3.8, and (Wessing and Eichelberg, 1979, Denholm, 2013). Also, it would be significant to obtain 3D images of whole-fly MTs to identify how the localisation of this organ affects the overall functions of the fly in conditions where the MTs are shorter than normal. Related to this, an interesting question to answer using this technology would be if the MT positioning is affected in *GATAe* knockdown conditions (as these MTs resulted significantly shorted compared to the controls, Figure 4.2.5), and if these possible defects in MT positioning also contribute to the strong functional deficiencies observed in the results of previous chapters.

7.5 Results

CapaR>GFP flies were used for this purpose, as they specifically expressed cytosolic GFP in PCs (Terhzaz et al., 2012). The CUBIC treatment (Richardson and Lichtman, 2015) was performed as a clearing method in these specimens see 2.7.2 and Appendix 8). After CUBIC clearing, samples were placed in the OPT

microscope and scanned at different angles, covering all 360°. After image processing using Imaris Software, the 3D images were reconstructed, and indistinguishable GFP areas could be observed in different parts of the body, (Figure 7.5.1A). However, no specific GFP expression was observed in the MTs. After using different approaches to optimise the protocol to clear the samples, the dorsal carcass was carefully removed after the fixation step, just before CUBIC treatment. This allowed increasing the transparency of the sample and detection of the specific GFP expression in the MTs. After the scanning process, specific and robust GFP expression was observed in the MTs, improving the results of this technique (Figure 7.5.1B and C).

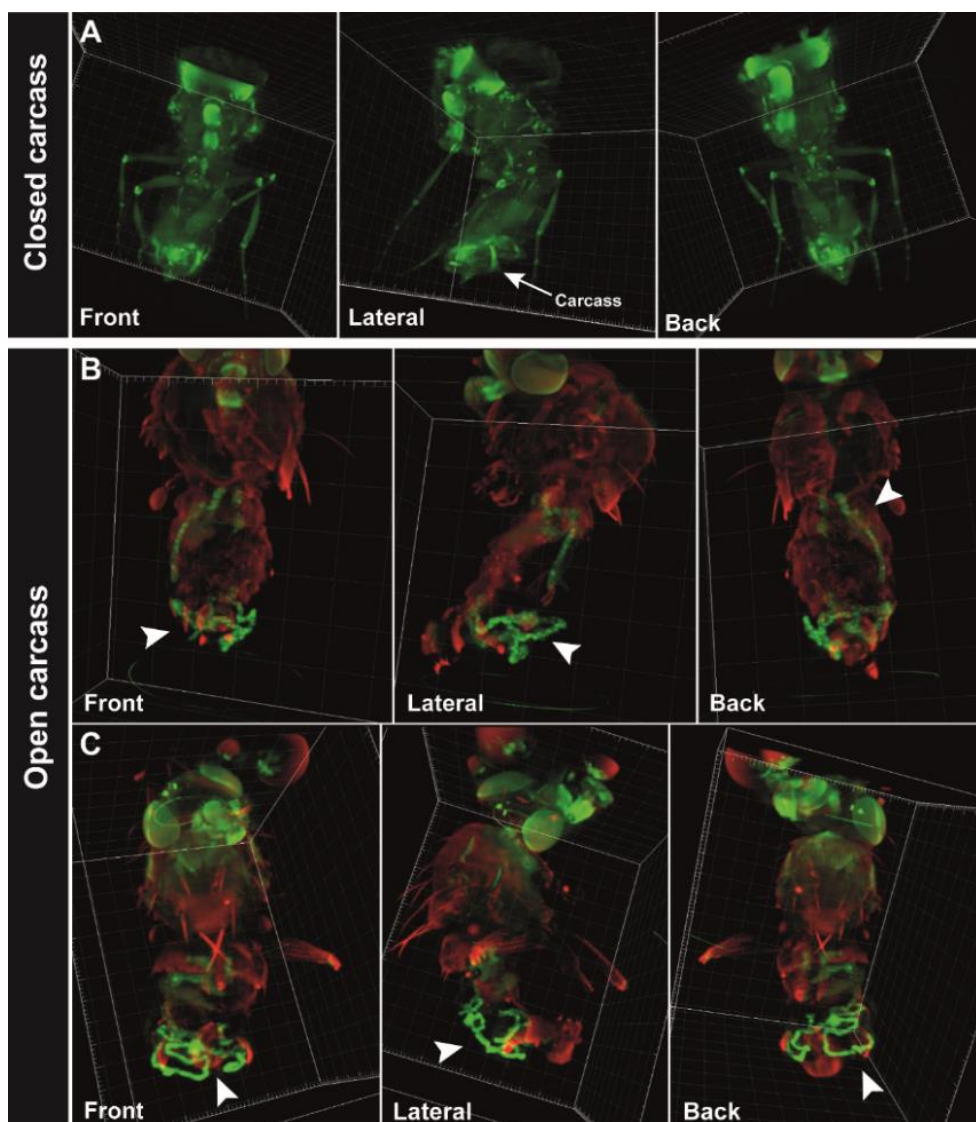


Figure 7.5.1. 3D images of whole adult flies using OPT microscopy.

Adult female flies treated with CUBIC for 24h. (A) *CapaR*>GFP flies in which the carcass has not been removed show nonspecific GFP expression all along the body, with no specific signal from the MTs. (B and C) Two examples of *CapaR*-GFP flies with the carcass removed before CUBIC treatment show strong GFP expression in their MTs (white arrows). GFP is shown in green, and bright field channel is shown in red.

To obtain higher resolution images of the adult MTs, specimens were scanned using a two-photon microscope. This microscopy technique allows the observation of 3D structures with higher sensitivity than confocal microscopy (Helmchen and Denk, 2005). This technique has been used to observe other tissues such as the MG or HG, and combined with label-free Third Harmonic Generation (THG) imaging (Weigelin et al., 2016). Combining the specific GFP signal and THG imaging, the MTs and surrounding tissues' morphology are easily observed in cellular detail in 3D (Figure 7.5.2). The resolution obtained using two-photon microscopy is considerably better than the resolution obtained using OPT, and, for example, the nuclei of the PCs can be easily distinguished (Figure 7.5.2B). Also, two-photon microscopy offers increased depth penetration compared to confocal microscopy, thereby allowing better 3D images of whole-fly MTs.

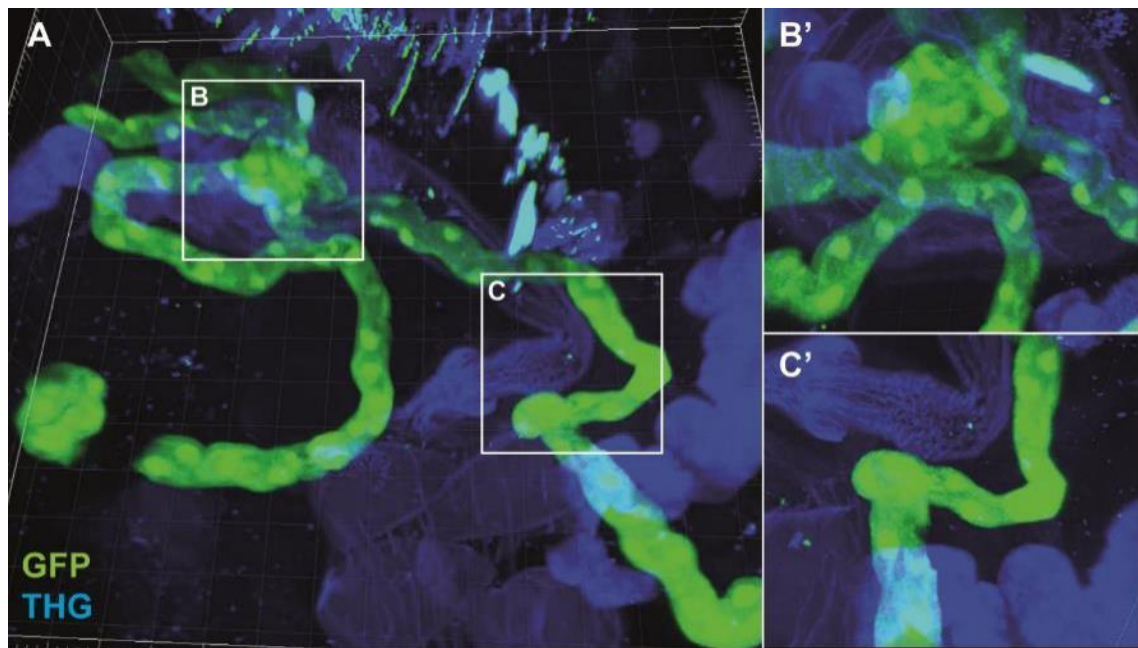


Figure 7.5.2. Adult MTs observed using multiphoton fluorescence microscopy.

(A) Adult posterior *CapaR*>GFP MTs in an intact fly scanned with a two-photon microscope. Single cells can be identified. B' and C' are zoomed projections of the B and C squares, respectively. C' shows the area of the hindgut visceral muscle where the MTs are attached (Hoch et al., 1994). All images are frames from a video generated with Imaris software. As these images are snapshots of the resulting .AVI file, they do not include a scale bar.

Embryo imaging was also performed using two-photon microscopy. Embryos were dechorionated and fixed, but not processed with a CUBIC solution, as they are natively transparent enough to be observed with no need of clearing protocols. The 3D structure of their MTs stained with ct antibody could be easily identified (Figure 7.5.3).

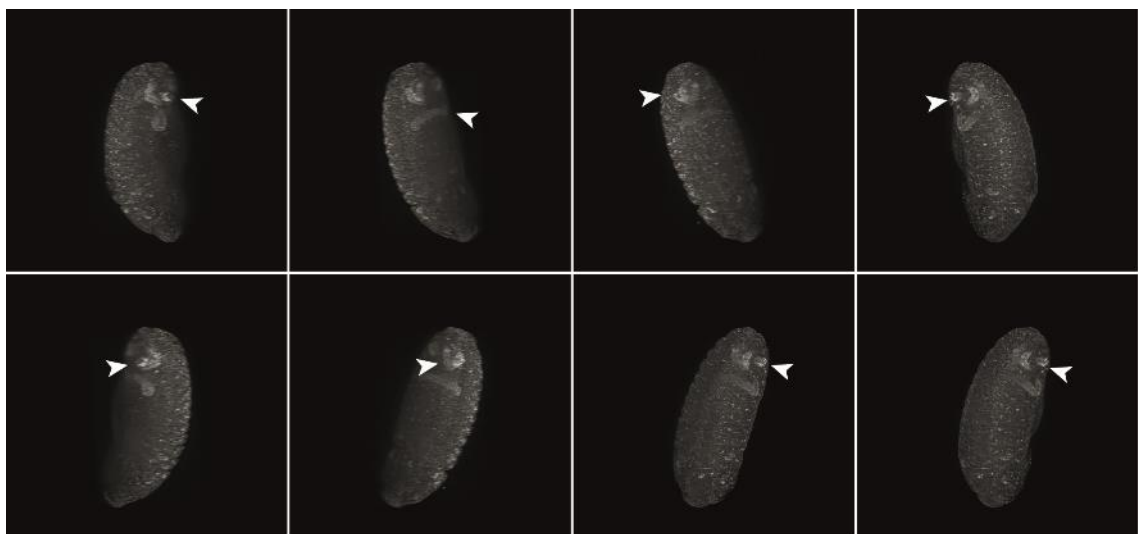


Figure 7.5.3. 3D structure of embryos observed using two-photon microscopy.

3D stacks of a wild-type embryo (~stage 10) immunostained with ct antibody. The structure of the MTs can be observed in all the stacks (arrowheads). Orientation: top is posterior and bottom anterior. As these images are snapshots of the resulting .AVI file, they do not include a scale bar.

7.6 Discussion

These experiments performed in BioCenter Oulu provided imaging of 3D intact organs without the need of dissections. Whole-specimen imaging methods are especially helpful for analyses where the localisation of the organ is essential, e.g. to identify defects in the migration of the MTs that could affect their stereotypical attachments.

OPT technology has been previously used in other systems, such as mouse, *Xenopus* or zebrafish (Susaki and Ueda, 2016), but it has not been yet used in invertebrates. One of the main problems of this technique in arthropods is the difficulty to permeabilize and clear the insect cuticle. In this study, it has been shown that *Drosophila* is suitable for this technology, as long as the cuticle is manually removed.

Alternative fixation systems could make the epidermis more permeable or transparent, e.g. via methanol fixation. This method of fixation can be recommended for specific types of tissues or antibodies, but the main disadvantages are that it does not preserve lipid structures or chromosome structures (Bonaccorsi et al., 2011). Another protocol step which can be modified is the clearing solution or the length of incubation. In this study, only the CUBCC solution has been used, but different methods, such as BABB or CLARITY (Susaki and Ueda, 2016) have been utilised in other systems that could also be suitable in *Drosophila* (Susaki and Ueda, 2016, Tainaka et al., 2016). A recent publication also presented an effective way to overcome some of these clearing difficulties (McConnell and Amos, 2018). Alternatively, other studies have investigated a non-invasive technology called X-ray micro-computed tomography (micro-CT) to stain whole-body *Drosophila* and internal organs (Fanenbruck et al., 2001), and recent work has used this same technology to stain the nervous system of this animal (Sombke et al., 2015), and a onychophoran limb (Müller et al., 2017) with stunning resolution results. Even though micro-CT technology allows a very high resolution and precise identification of internal organs, it is confined to small specimens, which can be a limiting condition for use in other organisms apart from *Drosophila*.

In contrast to the larval and adult samples, *Drosophila* embryos are easily imaged by two-photon microscopy, as they are more transparent and smaller. This allows an easier observation of 3D structures and tissue organisation in the embryonic development. However, although the images are shown here show a clear 3D structure of the MTs of the embryos, the results may not be much different using a standard confocal microscope. Therefore, the potential advantages that this technique could have using adult or larval specimens, it would not be applicable to embryonic samples, as, in the embryos, the use of a standard confocal microscope would be enough to observe any internal structures. However, it still would be a good for imaging live embryos, due to the improvement in depth and penetration.

In conclusion, the method presented in this chapter could serve as a complementary way to observe 3D structures in *Drosophila* and could be used in a similar way to the 3D imaging methods employed in vertebrate systems. Further experimentation with different protocols and clearing/fixating approaches would be required to refine and optimise this technique in the larval and adult stages of *Drosophila*.

Appendix 12 Publications

MARTÍNEZ-CORRALES, G., CABRERO, P., DOW, J. A., TERHZA, S. & DAVIES, S.-A. 2019. *Novel roles for GATAe in growth, maintenance and proliferation of cell populations in the Drosophila renal tubule.* **Development**, 146, dev178087.

DAVIES, S.-A., CABRERO, P., MARLEY, R., CORRALES, G. M., GHIMIRE, S., DORNAN, A. J. & DOW, J. A. 2019. *Epithelial Function in the Drosophila Malpighian Tubule: An In Vivo Renal Model.* **Kidney Organogenesis. Springer.**

References

- ACKERMAN, K. G., WANG, J., LUO, L., FUJIWARA, Y., ORKIN, S. H. & BEIER, D. R. 2007. Gata4 is necessary for normal pulmonary lobar development. *American journal of respiratory cell and molecular biology*, 36, 391-397.
- ADAMS, M. D., CELNIKER, S. E., HOLT, R. A., EVANS, C. A., GOCAYNE, J. D., AMANATIDES, P. G., SCHERER, S. E., LI, P. W., HOSKINS, R. A. & GALLE, R. F. 2000. The genome sequence of *Drosophila melanogaster*. *Science*, 287, 2185-2195.
- AGNIHOTRI, S., WOLF, A., PICARD, D., HAWKINS, C. & GUHA, A. 2009. GATA4 is a regulator of astrocyte cell proliferation and apoptosis in the human and murine central nervous system. *Oncogene*, 28, 3033.
- AINSWORTH, C., WAN, S. & SKAER, H. 2000. Coordinating cell fate and morphogenesis in *Drosophila* renal tubules. *Philos Trans R Soc Lond B Biol Sci*, 355, 931-7.
- AKDEMIR, F., CHRISTICH, A., SOGAME, N., CHAPO, J. & ABRAMS, J. 2007. p53 directs focused genomic responses in *Drosophila*. *Oncogene*, 26, 5184.
- AKIYAMA, Y., WATKINS, N., SUZUKI, H., JAIR, K.-W., VAN ENGELAND, M., ESTELLER, M., SAKAI, H., REN, C.-Y., YUASA, Y. & HERMAN, J. G. 2003. GATA-4 and GATA-5 transcription factor genes and potential downstream antitumor target genes are epigenetically silenced in colorectal and gastric cancer. *Molecular and Cellular Biology*, 23, 8429-8439.
- ALI, A., CHRISTIE, P. T., GRIGORIEVA, I. V., HARDING, B., VAN ESCH, H., AHMED, S. F., BITNER-GLINDZICZ, M., BLIND, E., BLOCH, C. & CHRISTIN, P. 2007. Functional characterization of GATA3 mutations causing the hypoparathyroidism-deafness-renal (HDR) dysplasia syndrome: insight into mechanisms of DNA binding by the GATA3 transcription factor. *Human molecular genetics*, 16, 265-275.
- ALIC, N., ANDREWS, T. D., GIANNAKOU, M. E., PAPATHEODOROU, I., SLACK, C., HODDINOTT, M. P., COCHEMÉ, H. M., SCHUSTER, E. F., THORNTON, J. M. & PARTRIDGE, L. 2011. Genome-wide dFOXO targets and topology of the transcriptomic response to stress and insulin signalling. *Molecular systems biology*, 7, 502.
- ALLAN, A. K., DU, J., DAVIES, S. A. & DOW, J. A. 2005. Genome-wide survey of V-ATPase genes in *Drosophila* reveals a conserved renal phenotype for lethal alleles. *Physiological genomics*, 22, 128-138.
- ARAÚJO, S. J., ASLAM, H., TEAR, G. & CASANOVA, J. 2005. mummy/cystic encodes an enzyme required for chitin and glycan synthesis, involved in trachea, embryonic cuticle and CNS development—analysis of its role in *Drosophila* tracheal morphogenesis. *Developmental biology*, 288, 179-193.
- ARAVIN, A. A., HANNON, G. J. & BRENNECKE, J. 2007. The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. *science*, 318, 761-764.

- ARIES, A., PARADIS, P., LEFEBVRE, C., SCHWARTZ, R. J. & NEMER, M. 2004. Essential role of GATA-4 in cell survival and drug-induced cardiotoxicity. *Proceedings of the National Academy of Sciences*, 101, 6975-6980.
- ARTERO, R. D., MONFERRER, L., GARCIA-LOPEZ, A. & BAYLIES, M. K. 2006. Serpent and a hbris reporter are co-expressed in migrating cells during *Drosophila* hematopoiesis and Malpighian tubule formation. *Hereditas*, 143, 117-22.
- ASHBURNER, M. & DRYSDALE, R. 1994. FlyBase—the *Drosophila* genetic database. *Development*, 120, 2077-2079.
- AUGHEY, G. N. & SOUTHALL, T. D. 2016. Dam it's good! DamID profiling of protein-DNA interactions. *Wiley Interdisciplinary Reviews: Developmental Biology*, 5, 25-37.
- AZPIAZU, N. & FRASCH, M. 1993. tinman and bagpipe: two homeo box genes that determine cell fates in the dorsal mesoderm of *Drosophila*. *Genes & Development*, 7, 1325-1340.
- BAECHLER, B. L., MCKNIGHT, C., PRUCHNICKI, P. C., BIRO, N. A. & REED, B. H. 2016. Hindsight/RREB-1 functions in both the specification and differentiation of stem cells in the adult midgut of *Drosophila*. *Biology open*, 5, 1-10.
- BAEHRECKE, E. 2000. Steroid regulation of programmed cell death during *Drosophila* development. *Cell death and differentiation*, 7, 1057.
- BARTMAN, A. E., BUISINE, M. P., AUBERT, J. P., NIEHANS, G. A., TORIBARA, N. W., KIM, Y. S., KELLY, E. J., CRABTREE, J. E. & HO, S. B. 1998. The MUC6 secretory mucin gene is expressed in a wide variety of epithelial tissues. *The Journal of Pathology: A Journal of the Pathological Society of Great Britain and Ireland*, 186, 398-405.
- BASSETT, A. R., TIBBIT, C., PONTING, C. P. & LIU, J.-L. 2013. Highly efficient targeted mutagenesis of *Drosophila* with the CRISPR/Cas9 system. *Cell reports*, 4, 220-228.
- BEAVEN, R. & DENHOLM, B. 2018. Release and spread of Wingless is required to pattern the proximo-distal axis of *Drosophila* renal tubules. *eLife*, 7, e35373.
- BEIRA, J. V. & PARO, R. 2016. The legacy of *Drosophila* imaginal discs. *Chromosoma*, 125, 573-592.
- BEJSOVEC, A. & WIESCHAUS, E. 1993. Segment polarity gene interactions modulate epidermal patterning in *Drosophila* embryos. *Development*, 119, 501-517.
- BELAGULI, N. S., AFTAB, M., RIGI, M., ZHANG, M., ALBO, D. & BERGER, D. H. 2010. GATA6 promotes colon cancer cell invasion by regulating urokinase plasminogen activator gene expression. *Neoplasia*, 12, 856-865.
- BELLEN, H. J., LEVIS, R. W., LIAO, G., HE, Y., CARLSON, J. W., TSANG, G., EVANS-HOLM, M., HIESINGER, P. R., SCHULZE, K. L. & RUBIN, G. M. 2004. The BDGP

gene disruption project: single transposon insertions associated with 40% of *Drosophila* genes. *Genetics*, 167, 761-781.

- BENJAMINI, Y., KRIEGER, A. M. & YEKUTIELI, D. 2006. Adaptive linear step-up procedures that control the false discovery rate. *Biometrika*, 93, 491-507.
- BERTOLONI MELI, D. 2011. *Mechanism, experiment, disease: Marcello Malpighi and seventeenth-century anatomy*, The Johns Hopkins University Press.
- BEULING, E., BAFFOUR-AWUAH, N. Y. A., STAPLETON, K. A., ARONSON, B. E., NOAH, T. K., SHROYER, N. F., DUNCAN, S. A., FLEET, J. C. & KRASINSKI, S. D. 2011. GATA factors regulate proliferation, differentiation, and gene expression in small intestine of mature mice. *Gastroenterology*, 140, 1219-1229. e2.
- BEYENBACH, K. W., SKAER, H. & DOW, J. A. 2010. The developmental, molecular, and transport biology of Malpighian tubules. *Annu Rev Entomol*, 55, 351-74.
- BIER, E. & BODMER, R. 2004. *Drosophila*, an emerging model for cardiac disease. *Gene*, 342, 1-11.
- BIER, E., VAESSIN, H., SHEPHERD, S., LEE, K., MCCALL, K., BARBEL, S., ACKERMAN, L., CARRETTO, R., UEMURA, T. & GRELL, E. 1989. Searching for pattern and mutation in the *Drosophila* genome with a P-lacZ vector. *Genes & development*, 3, 1273-1287.
- BLOCHLINGER, K., BODMER, R., JAN, L. Y. & JAN, Y. N. 1990. Patterns of expression of cut, a protein required for external sensory organ development in wild-type and cut mutant *Drosophila* embryos. *Genes & Development*, 4, 1322-1331.
- BODE, A. M. & DONG, Z. 2004. Post-translational modification of p53 in tumorigenesis. *Nature Reviews Cancer*, 4, 793.
- BOHÈRE, J., MANCHENO-FERRIS, A., AL HAYEK, S., ZANET, J., VALENTI, P., AKINO, K., YAMABE, Y., INAGAKI, S., CHANUT-DELALANDE, H., PLAZA, S., KAGEYAMA, Y., OSMAN, D., POLESELLO, C. & PAYRE, F. 2018. Shavenbaby and Yorkie mediate Hippo signaling to protect adult stem cells from apoptosis. *Nature Communications*, 9, 5123.
- BONACCORSI, S., GIANSAINTI, M. G., CENCI, G. & GATTI, M. 2011. Methanol-acetone fixation of *Drosophila* testes. *Cold Spring Harbor protocols*, 2011, pdb. prot065763.
- BORATYN, G. M., CAMACHO, C., COOPER, P. S., COULOURIS, G., FONG, A., MA, N., MADDEN, T. L., MATTEN, W. T., MCGINNIS, S. D. & MERZHUH, Y. 2013. BLAST: a more efficient report with usability improvements. *Nucleic acids research*, 41, W29-W33.
- BORTZ, P. D. S. & WAMHOFF, B. R. 2011. Chromatin immunoprecipitation (ChIP): revisiting the efficacy of sample preparation, sonication, quantification of sheared DNA, and analysis via PCR. *PloS one*, 6, e26015.

- BOULEAU, S. & TRICOIRE, H. 2015. Drosophila models of Alzheimer's disease: advances, limits, and perspectives. *Journal of Alzheimer's Disease*, 45, 1015-1038.
- BOYLE, E. I., WENG, S., GOLLUB, J., JIN, H., BOTSTEIN, D., CHERRY, J. M. & SHERLOCK, G. 2004. GO:: TermFinder—open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes. *Bioinformatics*, 20, 3710-3715.
- BRAND, A. H. & PERRIMON, N. 1993. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development*, 118, 401-15.
- BRESNICK, E. H., KATSUMURA, K. R., LEE, H.-Y., JOHNSON, K. D. & PERKINS, A. S. 2012. Master regulatory GATA transcription factors: mechanistic principles and emerging links to hematologic malignancies. *Nucleic acids research*, 40, 5819-5831.
- BROWN, K. 2017. An interview with Christiane Nüsslein-Volhard. *Development*, 144, 3851-3854.
- BROWN, S. & HOMBRIA, J. C.-G. 2000. Drosophila grain encodes a GATA transcription factor required for cell rearrangement during morphogenesis. *Development*, 127, 4867-4876.
- BROWNE, A. & O'DONNELL, M. J. 2013. Ammonium secretion by Malpighian tubules of *Drosophila melanogaster*: application of a novel ammonium-selective microelectrode. *Journal of Experimental Biology*, jeb. 091082.
- BRUNNER, E., PETER, O., SCHWEIZER, L. & BASLER, K. 1997. pangolin encodes a Lef-1 homologue that acts downstream of Armadillo to transduce the Wingless signal in *Drosophila*. *Nature*, 385, 829.
- BUCHON, N., OSMAN, D., DAVID, F. P., FANG, H. Y., BOQUETE, J. P., DEPLANCHE, B. & LEMAITRE, B. 2013. Morphological and molecular characterization of adult midgut compartmentalization in *Drosophila*. *Cell Rep*, 3, 1725-38.
- BUNT, S., HOOLEY, C., HU, N., SCAHILL, C., WEAVERS, H. & SKAER, H. 2010. Hemocyte-secreted type IV collagen enhances BMP signaling to guide renal tubule morphogenesis in *Drosophila*. *Developmental cell*, 19, 296-306.
- BÜRGLIN, T. R. & AFFOLTER, M. 2016. Homeodomain proteins: an update. *Chromosoma*, 125, 497-521.
- CABRERO, P., TERHIAZ, S., ROMERO, M. F., DAVIES, S. A., BLUMENTHAL, E. M. & DOW, J. A. 2014. Chloride channels in stellate cells are essential for uniquely high secretion rates in neuropeptide-stimulated *Drosophila* diuresis. *Proc Natl Acad Sci U S A*, 111, 14301-6.
- CAMPBELL, K., CASANOVA, J. & SKAER, H. 2010. Mesenchymal-to-epithelial transition of intercalating cells in *Drosophila* renal tubules depends on polarity cues from epithelial neighbours. *Mechanisms of development*, 127, 345-357.

- CAMPBELL, K., KNUST, E. & SKAER, H. 2009. Crumbs stabilises epithelial polarity during tissue remodelling. *Journal of cell science*, 122, 2604-2612.
- CAMPBELL, K., WHISSELL, G., FRANCH-MARRO, X., BATLLE, E. & CASANOVA, J. 2011. Specific GATA factors act as conserved inducers of an endodermal-EMT. *Developmental cell*, 21, 1051-1061.
- CARE, A., CATALUCCI, D., FELICETTI, F., BONCI, D., ADDARIO, A., GALLO, P., BANG, M.-L., SEGNALINI, P., GU, Y. & DALTON, N. D. 2007. MicroRNA-133 controls cardiac hypertrophy. *Nature medicine*, 13, 613.
- CAYGILL, E. E. & BRAND, A. H. 2016. The GAL4 System: A Versatile System for the Manipulation and Analysis of Gene Expression. *Drosophila*. Springer.
- CHANG, S.-K., DOHRMAN, A. F., BASBAUM, C. B., HO, S. B., TSUDA, T., TORIBARA, N. W., GUM, J. R. & KIM, Y. S. 1994. Localization of mucin (MUC2 and MUC3) messenger RNA and peptide expression in human normal intestine and colon cancer. *Gastroenterology*, 107, 28-36.
- CHANG, S., BRAY, S. M., LI, Z., ZARNESCU, D. C., HE, C., JIN, P. & WARREN, S. T. 2008. Identification of small molecules rescuing fragile X syndrome phenotypes in Drosophila. *Nature chemical biology*, 4, 256.
- CHI, T., KIM, M. S., LANG, S., BOSE, N., KAHN, A., FLECHNER, L., BLASCHKO, S. D., ZEE, T., MUTELIEFU, G. & BOND, N. 2015. A Drosophila model identifies a critical role for zinc in mineralization for kidney stone disease. *PloS one*, 10, e0124150.
- CHIA, I., GROTE, D., MARCOTTE, M., BATOURINA, E., MENDELSON, C. & BOUCHARD, M. 2011. Nephric duct insertion is a crucial step in urinary tract maturation that is regulated by a Gata3-Raldh2-Ret molecular network in mice. *Development*, 138, 2089-2097.
- CHIA, N.-Y., DENG, N., DAS, K., HUANG, D., HU, L., ZHU, Y., LIM, K. H., LEE, M.-H., WU, J. & SAM, X. X. 2014. Regulatory crosstalk between lineage-survival oncogenes KLF5, GATA4 and GATA6 cooperatively promotes gastric cancer development. *Gut*, gutjnl-2013-306596.
- CHINTAPALLI, V. R., TERHSAZ, S., WANG, J., AL BRATTY, M., WATSON, D. G., HERZYK, P., DAVIES, S. A. & DOW, J. A. 2012. Functional correlates of positional and gender-specific renal asymmetry in Drosophila. *PloS one*, 7, e32577.
- CHINTAPALLI, V. R., WANG, J. & DOW, J. A. 2007. Using FlyAtlas to identify better Drosophila melanogaster models of human disease. *Nat Genet*, 39, 715-20.
- CHOI, N. H., KIM, J. G., YANG, D. J., KIM, Y. S. & YOO, M. A. 2008. Age-related changes in Drosophila midgut are associated with PVF2, a PDGF/VEGF-like growth factor. *Aging cell*, 7, 318-334.
- COLLAS, P. 2010. The current state of chromatin immunoprecipitation. *Molecular biotechnology*, 45, 87-100.

- COLUSSI, P. A., QUINN, L. M., HUANG, D. C., COOMBE, M., READ, S. H., RICHARDSON, H. & KUMAR, S. 2000. Debcl, a proapoptotic Bcl-2 homologue, is a component of the *Drosophila melanogaster* cell death machinery. *J Cell Biol*, 148, 703-14.
- CONSORTIUM, D. G. 2007. Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature*, 450, 203.
- CONSORTIUM, G. O. 2004. The Gene Ontology (GO) database and informatics resource. *Nucleic acids research*, 32, D258-D261.
- COOPER, T. G. 2002. Transmitting the signal of excess nitrogen in *Saccharomyces cerevisiae* from the Tor proteins to the GATA factors: connecting the dots. *FEMS microbiology reviews*, 26, 223-238.
- DANESE, E., RUZZENENTE, A., MONTAGNANA, M. & LIEVENS, P. M.-J. 2018. Current and future roles of mucins in cholangiocarcinoma—recent evidences for a possible interplay with bile acids. *Annals of translational medicine*, 6.
- DAVIES, S.-A., CABRERO, P., MARLEY, R., CORRALES, G. M., GHIMIRE, S., DORNAN, A. J. & DOW, J. A. 2019. Epithelial Function in the *Drosophila* Malpighian Tubule: An In Vivo Renal Model. *Kidney Organogenesis*. Springer.
- DAVIES, S. A., CABRERO, P., OVEREND, G., AITCHISON, L., SEBASTIAN, S., TERHZAZ, S. & DOW, J. A. 2014a. Cell signalling mechanisms for insect stress tolerance. *J Exp Biol*, 217, 119-28.
- DAVIES, S. A., CABRERO, P., OVEREND, G., AITCHISON, L., SEBASTIAN, S., TERHZAZ, S. & DOW, J. A. T. 2014b. Cell signalling mechanisms for insect stress tolerance. *The Journal of Experimental Biology*, 217, 119-128.
- DAVIES, S. A., OVEREND, G., SEBASTIAN, S., CUNDALL, M., CABRERO, P., DOW, J. A. & TERHZAZ, S. 2012. Immune and stress response 'cross-talk' in the *Drosophila* Malpighian tubule. *J Insect Physiol*, 58, 488-97.
- DE CASTRO, E., SIGRIST, C. J., GATTIKER, A., BULLIARD, V., LANGENDIJK-GENEVAUX, P. S., GASTEIGER, E., BAIROCH, A. & HULO, N. 2006. ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic acids research*, 34, W362-W365.
- DE MADRID, B. H. & CASANOVA, J. 2018. GATA factor genes in the *Drosophila* midgut embryo. *PloS one*, 13, e0193612.
- DEAN, S., TANG, J. I. & SECKL, J. R. 2010. Developmental and tissue-specific regulation of hepatocyte nuclear factor 4- α (HNF4- α) isoforms in rodents. *Gene expression*, 14, 337-344.
- DENHOLM, B. 2013. Shaping up for action: the path to physiological maturation in the renal tubules of *Drosophila*. *Organogenesis*, 9, 40-54.
- DENHOLM, B., HU, N., FAUQUIER, T., CAUBIT, X., FASANO, L. & SKAER, H. 2013. The tiptop/teashirt genes regulate cell differentiation and renal physiology in *Drosophila*. *Development*, 140, 1100-1110.

- DENHOLM, B. & SKAER, H. 2009. Bringing together components of the fly renal system. *Curr Opin Genet Dev*, 19, 526-32.
- DENHOLM, B., SUDARSAN, V., PASALODOS-SANCHEZ, S., ARTERO, R., LAWRENCE, P., MADDRELL, S., BAYLIES, M. & SKAER, H. 2003. Dual origin of the renal tubules in *Drosophila*: mesodermal cells integrate and polarize to establish secretory function. *Curr Biol*, 13, 1052-7.
- DHANISHA, S. S., GURUVAYOORAPPAN, C., DRISHYA, S. & ABEESH, P. 2018. Mucins: Structural diversity, biosynthesis, its role in pathogenesis and as possible therapeutic targets. *Critical reviews in oncology/hematology*, 122, 98-122.
- DIETZL, G., CHEN, D., SCHNORRER, F., SU, K.-C., BARINOVA, Y., FELLNER, M., GASSER, B., KINSEY, K., OPPEL, S. & SCHEIBLAUER, S. 2007. A genome-wide transgenic RNAi library for conditional gene inactivation in *Drosophila*. *Nature*, 448, 151.
- DOBSON, A. J., HE, X., BLANC, E., BOLUKBASI, E., FESEHA, Y., YANG, M. & PIPER, M. D. 2018. Tissue-specific transcriptome profiling of *Drosophila* reveals roles for GATA transcription factors in longevity by dietary restriction. *npj Aging and Mechanisms of Disease*, 4, 5.
- DOW, J. A. 2003. The *Drosophila* phenotype gap—And how to close it. *Briefings in Functional Genomics*, 2, 121-127.
- DOW, J. A. 2007. Integrative physiology, functional genomics and the phenotype gap: a guide for comparative physiologists. *Journal of Experimental Biology*, 210, 1632-1640.
- DOW, J. A. 2012. The versatile stellate cell—more than just a space-filler. *Journal of insect physiology*, 58, 467-472.
- DOW, J. A. & DAVIES, S. A. 2003. Integrative physiology and functional genomics of epithelial function in a genetic model organism. *Physiological Reviews*, 83, 687-729.
- DOW, J. A. & DAVIES, S. A. 2006. The Malpighian tubule: rapid insights from post-genomic biology. *Journal of Insect Physiology*, 52, 365-378.
- DOW, J. A., MADDRELL, S. H., GORTZ, A., SKAER, N. J., BROGAN, S. & KAISER, K. 1994. The malpighian tubules of *Drosophila melanogaster*: a novel phenotype for studies of fluid secretion and its control. *J Exp Biol*, 197, 421-8.
- DOW, J. A. & ROMERO, M. F. 2010. *Drosophila* provides rapid modeling of renal development, function, and disease. *Am J Physiol Renal Physiol*, 299, F1237-44.
- DOZMOROV, M. G., HURST, R. E., CULKIN, D. J., KROPP, B. P., FRANK, M. B., OSBAN, J., PENNING, T. M. & LIN, H. K. 2009. Unique patterns of molecular profiling between human prostate cancer LNCaP and PC-3 cells. *The Prostate*, 69, 1077-1090.

- DRESSLER, G. R. & DOUGLASS, E. C. 1992. Pax-2 is a DNA-binding protein expressed in embryonic kidney and Wilms tumor. *Proceedings of the National Academy of Sciences*, 89, 1179-1183.
- DU, L., ZHOU, A., SOHR, A. & ROY, S. 2018. An Efficient Strategy for Generating Tissue-specific Binary Transcription Systems in *Drosophila* by Genome Editing. *JoVE (Journal of Visualized Experiments)*, e58268.
- DUBE, K., MCDONALD, D. & O'DONNELL, M. 2000. Calcium transport by isolated anterior and posterior Malpighian tubules of *Drosophila melanogaster*: roles of sequestration and secretion. *Journal of insect physiology*, 46, 1449-1460.
- DUNCAN, S. A., MANOVA, K., CHEN, W. S., HOODLESS, P., WEINSTEIN, D. C., BACHVAROVA, R. F. & DARNELL, J. 1994. Expression of transcription factor HNF-4 in the extraembryonic endoderm, gut, and nephrogenic tissue of the developing mouse embryo: HNF-4 is a marker for primary endoderm in the implanting blastocyst. *Proceedings of the National Academy of Sciences*, 91, 7598-7602.
- DUTTA, D., DOBSON, A. J., HOUTZ, P. L., GLASSER, C., REVAH, J., KORZELIUS, J., PATEL, P. H., EDGAR, B. A. & BUCHON, N. 2015. Regional Cell-Specific Transcriptome Mapping Reveals Regulatory Complexity in the Adult *Drosophila* Midgut. *Cell Rep*, 12, 346-58.
- EBERT, M. S. & SHARP, P. A. 2012. Roles for microRNAs in conferring robustness to biological processes. *Cell*, 149, 515-524.
- ELLIS, C. L., CHANG, A. G., CIMINO-MATHEWS, A., ARGANI, P., YOUSSEF, R. F., KAPUR, P., MONTGOMERY, E. A. & EPSTEIN, J. I. 2013. GATA-3 immunohistochemistry in the differential diagnosis of adenocarcinoma of the urinary bladder. *Am J Surg Pathol*, 37, 1756-60.
- FAHRBACH, S. E., SMAGGHE, G. & VELARDE, R. A. 2012. Insect nuclear receptors. *Annual review of entomology*, 57, 83-106.
- FANENBRUCK, M., DE CARLO, F. & MANCINI, D. 2001. Evaluating the advantage of X-ray microtomography in microanatomical studies of small arthropods. *APS Activity Reports. Argonne, IL: Argonne National Laboratory*.
- FEANY, M. B. & BENDER, W. W. 2000. A *Drosophila* model of Parkinson's disease. *Nature*, 404, 394.
- FEHON, R. G., KOOH, P. J., REBAY, I., REGAN, C. L., XU, T., MUSKAVITCH, M. A. & ARTAVANIS-TSAKONAS, S. 1990. Molecular interactions between the protein products of the neurogenic loci Notch and Delta, two EGF-homologous genes in *Drosophila*. *Cell*, 61, 523-534.
- FEI, P., YU, Z., WANG, X., LU, P. J., FU, Y., HE, Z., XIONG, J. & HUANG, Y. 2012. High dynamic range optical projection tomography (HDR-OPT). *Optics Express*, 20, 8824-8836.
- FERRARIS, S., DEL MONACO, A. G., GARELLI, E., CARANDO, A., DE VITO, B., PAPPI, P., LALA, R. & PONZONE, A. 2009. HDR syndrome: a novel "de novo" mutation in GATA3 gene. *Am J Med Genet A*, 149A, 770-5.

- FOSSETT, N., HYMAN, K., GAJEWSKI, K., ORKIN, S. H. & SCHULZ, R. A. 2003. Combinatorial interactions of serpent, lozenge, and U-shaped regulate crystal cell lineage commitment during *Drosophila* hematopoiesis. *Proceedings of the National Academy of Sciences*, 100, 11451-11456.
- FURLONG, E. E., ANDERSEN, E. C., NULL, B., WHITE, K. P. & SCOTT, M. P. 2001. Patterns of gene expression during *Drosophila* mesoderm development. *Science*, 293, 1629-1633.
- GAMBERI, C., HIPFNER, D. R., TRUDEL, M. & LUBELL, W. D. 2017. Bicaudal C mutation causes myc and TOR pathway up-regulation and polycystic kidney disease-like phenotypes in *Drosophila*. *PLoS genetics*, 13, e1006694.
- GARCÍA-BELLIDO, A. 2009. The cellular and genetic bases of organ size and shape in *Drosophila*. *International Journal of Developmental Biology*, 53, 1291-1303.
- GASPAR, P., HOLDER, M. V., AERNE, B. L., JANODY, F. & TAPON, N. 2015. Zyxin antagonizes the FERM protein expanded to couple F-actin and Yorkie-dependent organ growth. *Current biology*, 25, 679-689.
- GASTEIGER, E., HOOGLAND, C., GATTIKER, A., WILKINS, M. R., APPEL, R. D. & BAIROCH, A. 2005. Protein identification and analysis tools on the ExPASy server. *The proteomics protocols handbook*. Springer.
- GAUL, U. & WEIGEL, D. 1990. Regulation of Krüppel expression in the anlage of the Malpighian tubules in the *Drosophila* embryo. *Mechanisms of development*, 33, 57-67.
- GAUTAM, N. K. & TAPADIA, M. G. 2010. Ecdysone signaling is required for proper organization and fluid secretion of stellate cells in the Malpighian tubules of *Drosophila melanogaster*. *Int J Dev Biol*, 54, 635-42.
- GAUTAM, N. K., VERMA, P. & TAPADIA, M. G. 2015. Ecdysone regulates morphogenesis and function of Malpighian tubules in *Drosophila melanogaster* through EcR-B2 isoform. *Dev Biol*, 398, 163-76.
- GEORGIEV, G. P. 1984. Mobile genetic elements in animal cells and their biological significance. *The FEBS Journal*, 145, 203-220.
- GHAJAR, C. M., PEINADO, H., MORI, H., MATEI, I. R., EVASON, K. J., BRAZIER, H., ALMEIDA, D., KOLLER, A., HAJJAR, K. A. & STAINIER, D. Y. 2013. The perivascular niche regulates breast tumour dormancy. *Nature cell biology*, 15, 807.
- GHAVI-HELM, Y., ZHAO, B. & FURLONG, E. E. 2016. Chromatin immunoprecipitation for analyzing transcription factor binding and histone modifications in *Drosophila*. *Drosophila*. Springer.
- GILCHRIST, D. A., FARGO, D. C. & ADELMAN, K. 2009. Using ChIP-chip and ChIP-seq to study the regulation of gene expression: genome-wide localization studies reveal widespread regulation of transcription elongation. *Methods*, 48, 398-408.

- GILLIS, W. Q., BOWERMAN, B. A. & SCHNEIDER, S. Q. 2008. The evolution of protostome GATA factors: molecular phylogenetics, synteny, and intron/exon structure reveal orthologous relationships. *BMC evolutionary biology*, 8, 112.
- GRATZ, S. J., CUMMINGS, A. M., NGUYEN, J. N., HAMM, D. C., DONOHUE, L. K., HARRISON, M. M., WILDONGER, J. & O'CONNOR-GILES, K. M. 2013. Genome engineering of *Drosophila* with the CRISPR RNA-guided Cas9 nuclease. *Genetics*, genetics. 113.152710.
- GRATZ, S. J., RUBINSTEIN, C. D., HARRISON, M. M., WILDONGER, J. & O'CONNOR-GILES, K. M. 2015. CRISPR-Cas9 genome editing in *Drosophila*. *Current protocols in molecular biology*, 111, 31.2. 1-31.2. 20.
- GREGORY, T., YU, C., MA, A., ORKIN, S. H., BLOBEL, G. A. & WEISS, M. J. 1999. GATA-1 and erythropoietin cooperate to promote erythroid cell survival by regulating bcl-xL expression. *Blood*, 94, 87-96.
- GROTE, D., SOUABNI, A., BUSSLINGER, M. & BOUCHARD, M. 2006. Pax2/8-regulated Gata3 expression is necessary for morphogenesis and guidance of the nephric duct in the developing kidney. *Development*, 133, 53-61.
- HAENLIN, M., CUBADDA, Y., BLONDEAU, F., HEITZLER, P., LUTZ, Y., SIMPSON, P. & RAMAIN, P. 1997. Transcriptional activity of pannier is regulated negatively by heterodimerization of the GATA DNA-binding domain with a cofactor encoded by the u-shaped gene of *Drosophila*. *Genes & development*, 11, 3096-3108.
- HALBERG, K. A., RAINEY, S. M., VELAND, I. R., NEUERT, H., DORNAN, A. J., KLAMBT, C., DAVIES, S. A. & DOW, J. A. 2016. The cell adhesion molecule Fasciclin2 regulates brush border length and organization in *Drosophila* renal tubules. *Nat Commun*, 7, 11266.
- HALBERG, K. A., TERHIAZ, S., CABRERO, P., DAVIES, S. A. & DOW, J. A. 2015. Tracing the evolutionary origins of insect renal function. *Nat Commun*, 6, 6800.
- HATTON-ELLIS, E., AINSWORTH, C., SUSHAMA, Y., WAN, S., VIJAYRAGHAVAN, K. & SKAER, H. 2007. Genetic regulation of patterned tubular branching in *Drosophila*. *Proc Natl Acad Sci U S A*, 104, 169-74.
- HAVERI, H., WESTERHOLM-ORMIO, M., LINDFORS, K., MÄKI, M., SAVILAHTI, E., ANDERSSON, L. C. & HEIKINHEIMO, M. 2008. Transcription factors GATA-4 and GATA-6 in normal and neoplastic human gastrointestinal mucosa. *BMC gastroenterology*, 8, 9.
- HAYES, S. A., MILLER, J. M. & HOSHIZAKI, D. K. 2001. serpent, a GATA-like transcription factor gene, induces fat-cell development in *Drosophila melanogaster*. *Development*, 128, 1193-1200.
- HE, C., CHENG, H. & ZHOU, R. 2007. GATA family of transcription factors of vertebrates: phylogenetics and chromosomal synteny. *Journal of biosciences*, 32, 1273-1280.

- HEIKINHEIMO, M., ERMOLAEVA, M., BIELINSKA, M., RAHMAN, N. A., NARITA, N., HUHTANIEMI, I. T., TAPANAINEN, J. S. & WILSON, D. B. 1997. Expression and hormonal regulation of transcription factors GATA-4 and GATA-6 in the mouse ovary. *Endocrinology*, 138, 3505-3514.
- HEINZ, S., BENNER, C., SPANN, N., BERTOLINO, E., LIN, Y. C., LASLO, P., CHENG, J. X., MURRE, C., SINGH, H. & GLASS, C. K. 2010. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular cell*, 38, 576-589.
- HELLEBREKERS, D. M., LENTJES, M. H., VAN DEN BOSCH, S. M., MELOTTE, V., WOUTERS, K. A., DAENEN, K. L., SMITS, K. M., AKIYAMA, Y., YUASA, Y. & SANDULEANU, S. 2009. GATA4 and GATA5 are potential tumor suppressors and biomarkers in colorectal cancer. *Clinical cancer research*, 15, 3990-3997.
- HELMCHEN, F. & DENK, W. 2005. Deep tissue two-photon microscopy. *Nature methods*, 2, 932.
- HIRATA, T., CABRERO, P., BERKHOLZ, D. S., BONDESON, D. P., RITMAN, E. L., THOMPSON, J. R., DOW, J. A. & ROMERO, M. F. 2012. In vivo *Drosophila* genetic model for calcium oxalate nephrolithiasis. *American Journal of Physiology-Renal Physiology*, 303, F1555-F1562.
- HOCH, M., BROADIE, K., JACKLE, H. & SKAER, H. 1994. Sequential fates in a single cell are established by the neurogenic cascade in the Malpighian tubules of *Drosophila*. *Development*, 120, 3439-50.
- HOSOYA, T., MAILLARD, I. & ENGEL, J. D. 2010. From the cradle to the grave: activities of GATA-3 throughout T-cell development and differentiation. *Immunological reviews*, 238, 110-125.
- HUANG, R.-C. 2018. The discoveries of molecular mechanisms for the circadian rhythm: The 2017 Nobel Prize in Physiology or Medicine. *Biomedical journal*, 41, 5-8.
- JACK, J. & DELOTTO, Y. 1995. Structure and regulation of a complex locus: the cut gene of *Drosophila*. *Genetics*, 139, 1689-1700.
- JACK, J. & MYETTE, G. 1999. Mutations that alter the morphology of the malpighian tubules in *Drosophila*. *Development genes and evolution*, 209, 546-554.
- JANEWAY JR, C. A., TRAVERS, P., WALPORT, M. & SHLOMCHIK, M. J. 2001. The structure of a typical antibody molecule.
- JIANG, J.-Q., LI, R.-G., WANG, J., LIU, X.-Y., XU, Y.-J., FANG, W.-Y., CHEN, X.-Z., ZHANG, W., WANG, X.-Z. & YANG, Y.-Q. 2013. Prevalence and spectrum of GATA5 mutations associated with congenital heart disease. *International journal of cardiology*, 165, 570-573.
- JONCKHEERE, N., VELGHE, A., DUCOUROUBLE, M. P., COPIN, M. C., RENES, I. B. & VAN SEUNINGEN, I. 2011. The mouse *Muc5b* mucin gene is

transcriptionally regulated by thyroid transcription factor-1 (TTF-1) and GATA-6 transcription factors. *FEBS J*, 278, 282-94.

- JUNG, A. C., DENHOLM, B., SKAER, H. & AFFOLTER, M. 2005. Renal tubule development in *Drosophila*: a closer look at the cellular level. *J Am Soc Nephrol*, 16, 322-8.
- KAMNASARAN, D., QIAN, B., HAWKINS, C., STANFORD, W. L. & GUHA, A. 2007. GATA6 is an astrocytoma tumor suppressor gene identified by gene trapping of mouse glioma model. *Proceedings of the National Academy of Sciences*, 104, 8053-8058.
- KANEHISA, M., GOTO, S., KAWASHIMA, S. & NAKAYA, A. 2002. The KEGG databases at GenomeNet. *Nucleic acids research*, 30, 42-46.
- KARUNARATNE, A., HARGRAVE, M., POH, A. & YAMADA, T. 2002. GATA proteins identify a novel ventral interneuron subclass in the developing chick spinal cord. *Developmental biology*, 249, 30-43.
- KAUFMANN, N., MATHAI, J. C., HILL, W. G., DOW, J. A., ZEIDEL, M. L. & BRODSKY, J. L. 2005. Developmental expression and biophysical characterization of a *Drosophila melanogaster* aquaporin. *American Journal of Physiology-Cell Physiology*, 289, C397-C407.
- KERBER, B., FELLERT, S. & HOCH, M. 1998. Seven-up, the *Drosophila* homolog of the COUP-TF orphan receptors, controls cell proliferation in the insect kidney. *Genes & development*, 12, 1781-1786.
- KERMAN, B. E., CHESHIRE, A. M., MYAT, M. M. & ANDREW, D. J. 2008. Ribbon modulates apical membrane during tube elongation through Crumbs and Moesin. *Developmental biology*, 320, 278-288.
- KING, B. & DENHOLM, B. 2014. Malpighian tubule development in the red flour beetle (*Tribolium castaneum*). *Arthropod structure & development*, 43, 605-613.
- KISPERT, A., HERRMANN, B. G., LEPTIN, M. & REUTER, R. 1994. Homologs of the mouse Brachyury gene are involved in the specification of posterior terminal structures in *Drosophila*, *Tribolium*, and *Locusta*. *Genes & development*, 8, 2137-2150.
- KLEINO, A., VALANNE, S., ULVILA, J., KALLIO, J., MYLLYMÄKI, H., ENWALD, H., STÖVEN, S., POIDEVIN, M., UEDA, R. & HULTMARK, D. 2005. Inhibitor of apoptosis 2 and TAK1-binding protein are components of the *Drosophila* Imd pathway. *The EMBO journal*, 24, 3423-3434.
- KLUG, A. & RHODES, D. Zinc fingers: a novel protein fold for nucleic acid recognition. Cold Spring Harbor symposia on quantitative biology, 1987. Cold Spring Harbor Laboratory Press, 473-482.
- KOBAYASHI, S., LACKEY, T., HUANG, Y., BISPING, E., PU, W. T., BOXER, L. M. & LIANG, Q. 2006. Transcription factor *gata4* regulates cardiac BCL2 gene expression in vitro and in vivo. *The FASEB journal*, 20, 800-802.

- KOBAYASHI, S., VOLDEN, P., TIMM, D., MAO, K., XU, X. & LIANG, Q. 2010. Transcription factor GATA4 inhibits doxorubicin-induced autophagy and cardiomyocyte death. *Journal of Biological Chemistry*, 285, 793-804.
- KODO, K., NISHIZAWA, T., FURUTANI, M., ARAI, S., YAMAMURA, E., JOO, K., TAKAHASHI, T., MATSUOKA, R. & YAMAGISHI, H. 2009. GATA6 mutations cause human cardiac outflow tract defects by disrupting semaphorin-plexin signaling. *Proceedings of the National Academy of Sciences*, 106, 13933-13938.
- KOUROS-MEHR, H., BECHIS, S. K., SLORACH, E. M., LITTLEPAGE, L. E., EGEBLAD, M., EWALD, A. J., PAI, S.-Y., HO, I.-C. & WERB, Z. 2008. GATA-3 links tumor differentiation and dissemination in a luminal breast cancer model. *Cancer cell*, 13, 141-152.
- KRISHNA, S. S., MAJUMDAR, I. & GRISHIN, N. V. 2003. Structural classification of zinc fingers: survey and summary. *Nucleic acids research*, 31, 532-550.
- KUO, C. T., MORRISEY, E. E., ANANDAPPA, R., SIGRIST, K., LU, M. M., PARMACEK, M. S., SOUDAIS, C. & LEIDEN, J. M. 1997. GATA4 transcription factor is required for ventral morphogenesis and heart tube formation. *Genes & development*, 11, 1048-1060.
- KVON, E. Z., KAZMAR, T., STAMPFEL, G., YÁÑEZ-CUNA, J. O., PAGANI, M., SCHERNHUBER, K., DICKSON, B. J. & STARK, A. 2014. Genome-scale functional characterization of Drosophila developmental enhancers in vivo. *Nature*, 512, 91.
- KWEI, K. A., BASHYAM, M. D., KAO, J., RATHEESH, R., REDDY, E. C., KIM, Y. H., MONTGOMERY, K., GIACOMINI, C. P., CHOI, Y.-L. & CHATTERJEE, S. 2008. Genomic profiling identifies GATA6 as a candidate oncogene amplified in pancreaticobiliary cancer. *PLoS genetics*, 4, e1000081.
- LAFOREST, B., ANDELFINGER, G. & NEMER, M. 2011. Loss of Gata5 in mice leads to bicuspid aortic valve. *The Journal of clinical investigation*, 121, 2876-2887.
- LAN, Q., CAO, M., KOLLIPARA, R. K., ROSA, J. B., KITTLER, R. & JIANG, H. 2018. FoxA transcription factor Fork head maintains the intestinal stem/progenitor cell identities in Drosophila. *Developmental biology*, 433, 324-343.
- LATCHMAN, D. S. 1997. Transcription factors: an overview. *The international journal of biochemistry & cell biology*, 29, 1305-1312.
- LAUGIER, E., YANG, Z., FASANO, L., KERRIDGE, S. & VOLA, C. 2005. A critical role of teashirt for patterning the ventral epidermis is masked by ectopic expression of tiptop, a paralog of teashirt in Drosophila. *Developmental biology*, 283, 446-458.
- LEADER, D. P., KRAUSE, S. A., PANDIT, A., DAVIES, S. A. & DOW, J. A. T. 2018. FlyAtlas 2: a new version of the Drosophila melanogaster expression atlas with RNA-Seq, miRNA-Seq and sex-specific data. *Nucleic Acids Res*, 46, D809-D815.

- LEE, T., FEIG, L. & MONTELL, D. J. 1996. Two distinct roles for Ras in a developmentally regulated cell migration. *Development*, 122, 409-418.
- LEE, T. & LUO, L. 1999. Mosaic analysis with a repressible cell marker for studies of gene function in neuronal morphogenesis. *Neuron*, 22, 451-461.
- LEE, T. I. & YOUNG, R. A. 2013. Transcriptional regulation and its misregulation in disease. *Cell*, 152, 1237-51.
- LENTJES, M. H., NIESSEN, H. E., AKIYAMA, Y., DE BRUINE, A. P., MELOTTE, V. & VAN ENGELAND, M. 2016. The emerging role of GATA transcription factors in development and disease. *Expert Rev Mol Med*, 18, e3.
- LI, Y., PANG, Z., HUANG, H., WANG, C., CAI, T. & XI, R. 2017. Transcription factor antagonism controls enteroendocrine cell specification from intestinal stem cells. *Scientific reports*, 7, 988.
- LI, Z., GODINHO, F. J., KLUSMANN, J.-H., GARRIGA-CANUT, M., YU, C. & ORKIN, S. H. 2005. Developmental stage-selective effect of somatically mutated leukemogenic transcription factor GATA1. *Nature genetics*, 37, 613.
- LI, Z., LIU, S. & CAI, Y. 2015. EGFR/MAPK signaling regulates the proliferation of *Drosophila* renal and nephric stem cells. *J Genet Genomics*, 42, 9-20.
- LINDEBOOM, F., GILLEMANS, N., KARIS, A., JAEGLE, M., MEIJER, D., GROSVELD, F. & PHILIPSEN, S. 2003. A tissue-specific knockout reveals that Gata1 is not essential for Sertoli cell function in the mouse. *Nucleic acids research*, 31, 5405-5412.
- LIU, J., GHANIM, M., XUE, L., BROWN, C. D., IOSSIFOV, I., ANGELETTI, C., HUA, S., NÈGRE, N., LUDWIG, M. & STRICKER, T. 2009. Analysis of *Drosophila* segmentation network identifies a JNK pathway factor overexpressed in kidney cancer. *Science*, 323, 1218-1222.
- LOURENÇO, D., BRAUNER, R., RYBCZYŃSKA, M., NIHOUL-FÉKÉTÉ, C., MCELREAVEY, K. & BASHAMBOO, A. 2011. Loss-of-function mutation in GATA4 causes anomalies of human testicular development. *Proceedings of the National Academy of Sciences*, 108, 1597-1602.
- LOVATO, T. L., SENSIBAUGH, C. A., SWINGLE, K. L., MARTINEZ, M. M. & CRIPPS, R. M. 2015. The *Drosophila* transcription factors Tinman and Pannier activate and collaborate with Myocyte enhancer factor-2 to promote heart cell fate. *PloS one*, 10, e0132965.
- LOWRY, J. A. & ATCHLEY, W. R. 2000. Molecular evolution of the GATA family of transcription factors: conservation within the DNA-binding domain. *Journal of molecular evolution*, 50, 103-115.
- MAEDER, M. L., LINDER, S. J., REYON, D., ANGSTMANN, J. F., FU, Y., SANDER, J. D. & JOUNG, J. K. 2013. Robust, synergistic regulation of human gene expression using TALE activators. *Nature methods*, 10, 243.
- MAGLICH, J. M., SLUDER, A., GUAN, X., SHI, Y., MCKEE, D. D., CARRICK, K., KAMDAR, K., WILLSON, T. M. & MOORE, J. T. 2001. Comparison of complete

nuclear receptor sets from the human, *Caenorhabditis elegans* and *Drosophila* genomes. *Genome biology*, 2, research0029. 1.

- MARCHLER-BAUER, A., BO, Y., HAN, L., HE, J., LANCZYCKI, C. J., LU, S., CHITSAZ, F., DERBYSHIRE, M. K., GEER, R. C. & GONZALES, N. R. 2016. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic acids research*, 45, D200-D203.
- MARSHALL, O. J., SOUTHALL, T. D., CHEETHAM, S. W. & BRAND, A. H. 2016. Cell-type-specific profiling of protein-DNA interactions without cell isolation using targeted DamID with next-generation sequencing. *Nature protocols*, 11, 1586.
- MARTOVETSKY, G., TEE, J. B. & NIGAM, S. K. 2013. Hepatocyte nuclear factors 4 α and 1 α regulate kidney developmental expression of drug-metabolizing enzymes and drug transporters. *Molecular pharmacology*, 84, 808-823.
- MARUYAMA, R., GREVENGOED, E., STEMPNIEWICZ, P. & ANDREW, D. J. 2011. Genome-wide analysis reveals a major role in cell fate maintenance and an unexpected role in endoreduplication for the *Drosophila* FoxA gene Fork head. *PloS one*, 6, e20901.
- MCCONNELL, G. & AMOS, W. B. 2018. Application of the Mesolens for subcellular resolution imaging of intact larval and whole adult *Drosophila*. *Journal of microscopy*, 270, 252-258.
- MCGETTIGAN, J., MCLENNAN, R., BRODERICK, K., KEAN, L., ALLAN, A., CABRERO, P., REGULSKI, M., POLLOCK, V., GOULD, G. & DAVIES, S.-A. 2005. Insect renal tubules constitute a cell-autonomous immune system that protects the organism against bacterial infection. *Insect biochemistry and molecular biology*, 35, 741-754.
- MCGUIRE, S. E., MAO, Z. & DAVIS, R. L. 2004. Spatiotemporal gene expression targeting with the TARGET and gene-switch systems in *Drosophila*. *Sci. STKE*, 2004, pl6-pl6.
- MEAD, P. E., DECONINCK, A. E., HUBER, T. L., ORKIN, S. H. & ZON, L. I. 2001. Primitive erythropoiesis in the *Xenopus* embryo: the synergistic role of LMO-2, SCL and GATA-binding proteins. *Development*, 128, 2301-2308.
- MEIRELES-FILHO, A. C., BARDET, A. F., YÁÑEZ-CUNA, J. O., STAMPFEL, G. & STARK, A. 2014. Cis-regulatory requirements for tissue-specific programs of the circadian clock. *Current Biology*, 24, 1-10.
- MENGHINI, R., MARCHETTI, V., CARDELLINI, M., HRIBAL, M. L., MAURIELLO, A., LAURO, D., SBRACCIA, P., LAURO, R. & FEDERICI, M. 2005. Phosphorylation of GATA2 by Akt increases adipose tissue differentiation and reduces adipose tissue-related inflammation: a novel pathway linking obesity to atherosclerosis. *Circulation*, 111, 1946-1953.
- MENORET, D., SANTOLINI, M., FERNANDES, I., SPOKONY, R., ZANET, J., GONZALEZ, I., LATAPIE, Y., FERRER, P., ROUAULT, H. & WHITE, K. P. 2013. Genome-wide analyses of Shavenbaby target genes reveals distinct features of enhancer organization. *Genome biology*, 14, R86.

- MERIKA, M. & ORKIN, S. H. 1993. DNA-binding specificity of GATA family transcription factors. *Molecular and cellular biology*, 13, 3999-4010.
- MI, H., HUANG, X., MURUGANUJAN, A., TANG, H., MILLS, C., KANG, D. & THOMAS, P. D. 2016. PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. *Nucleic acids research*, 45, D183-D189.
- MI, H., MURUGANUJAN, A., CASAGRANDE, J. T. & THOMAS, P. D. 2013. Large-scale gene function analysis with the PANTHER classification system. *Nature protocols*, 8, 1551.
- MICCHELLI, C. A. & PERRIMON, N. 2006. Evidence that stem cells reside in the adult *Drosophila* midgut epithelium. *Nature*, 439, 475.
- MILLET-BOUREIMA, C., PORRAS MARROQUIN, J. & GAMBERI, C. 2018. Modeling Renal Disease “On the Fly”. *BioMed Research International*, 2018.
- MOLKENTIN, J. D., TYMITZ, K. M., RICHARDSON, J. A. & OLSON, E. N. 2000. Abnormalities of the genitourinary tract in female mice lacking GATA5. *Molecular and Cellular Biology*, 20, 5256-5260.
- MONDAL, K., DASTIDAR, A. G., SINGH, G., MADHUSUDHANAN, S., GANDE, S. L., VIJAYRAGHAVAN, K. & VARADARAJAN, R. 2007. Design and isolation of temperature-sensitive mutants of Gal4 in yeast and *Drosophila*. *Journal of molecular biology*, 370, 939-950.
- MORGAN, T. H. 1910. Sex Limited Inheritance in *Drosophila*. *Science*, 32, 120-2.
- MORRISEY, E. E., IP, H. S., LU, M. M. & PARMACEK, M. S. 1996. GATA-6: a zinc finger transcription factor that is expressed in multiple cell lineages derived from lateral mesoderm. *Developmental biology*, 177, 309-322.
- MORRISEY, E. E., IP, H. S., TANG, Z. & PARMACEK, M. S. 1997. GATA-4 activates transcription via two novel domains that are conserved within the GATA-4/5/6 subfamily. *Journal of Biological Chemistry*, 272, 8515-8524.
- MORRISEY, E. E., TANG, Z., SIGRIST, K., LU, M. M., JIANG, F., IP, H. S. & PARMACEK, M. S. 1998. GATA6 regulates HNF4 and is required for differentiation of visceral endoderm in the mouse embryo. *Genes & development*, 12, 3579-3590.
- MUKHERJEE, S. & HANLEY, K. A. 2010. RNA interference modulates replication of dengue virus in *Drosophila melanogaster* cells. *BMC microbiology*, 10, 127.
- MÜLLER, M., DE SENA OLIVEIRA, I., ALLNER, S., FERSTL, S., BIDOLA, P., MECHLEM, K., FEHRINGER, A., HEHN, L., DIEROLF, M. & ACHTERHOLD, K. 2017. Myoanatomy of the velvet worm leg revealed by laboratory-based nanofocus X-ray source tomography. *Proceedings of the National Academy of Sciences*, 114, 12378-12383.
- MURAKAMI, R., OKUMURA, T. & UCHIYAMA, H. 2005. GATA factors as key regulatory molecules in the development of *Drosophila* endoderm. *Dev Growth Differ*, 47, 581-9.

- MURAKAMI, R., TAKASHIMA, S. & HAMAGUCHI, T. 1999. Developmental genetics of the *Drosophila* gut: specification of primordia, subdivision and overt-differentiation. *Cellular and molecular biology (Noisy-le-Grand, France)*, 45, 661-676.
- NAWIJN, M. C., FERREIRA, R., DINGJAN, G. M., KAHRE, O., DRABEK, D., KARIS, A., GROSVELD, F. & HENDRIKS, R. W. 2001. Enforced expression of GATA-3 during T cell development inhibits maturation of CD8 single-positive cells and induces thymic lymphoma in transgenic mice. *The Journal of Immunology*, 167, 715-723.
- NITTA, K. R., JOLMA, A., YIN, Y., MORGUNOVA, E., KIVIOJA, T., AKHTAR, J., HENS, K., TOIVONEN, J., DEPLANCHE, B., FURLONG, E. E. & TAIPALE, J. 2015. Conservation of transcription factor binding specificities across 600 million years of bilateria evolution. *Elife*, 4.
- NOGI, Y., SHIMADA, H., MATSUZAKI, Y., HASHIMOTO, H. & FUKASAWA, T. 1984. Regulation of expression of the galactose gene cluster in *Saccharomyces cerevisiae*. *Molecular and General Genetics MGG*, 195, 29-34.
- O'HARE, K. & RUBIN, G. M. 1983. Structures of P transposable elements and their sites of insertion and excision in the *Drosophila melanogaster* genome. *Cell*, 34, 25-35.
- O'KANE, C. J. & GEHRING, W. J. 1987. Detection in situ of genomic regulatory elements in *Drosophila*. *Proceedings of the National Academy of Sciences*, 84, 9123-9127.
- O'DONNELL, M. J., RHEAULT, M. R., DAVIES, S. A., ROSAY, P., HARVEY, B. J., MADDRELL, S. H., KAISER, K. & DOW, J. A. 1998. Hormonally controlled chloride movement across *Drosophila* tubules is via ion channels in stellate cells. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, 274, R1039-R1049.
- OH, H., SLATTERY, M., MA, L., CROFTS, A., WHITE, K. P., MANN, R. S. & IRVINE, K. D. 2013. Genome-wide association of Yorkie with chromatin and chromatin-remodeling complexes. *Cell reports*, 3, 309-318.
- OHLSTEIN, B. & SPRADLING, A. 2006. The adult *Drosophila* posterior midgut is maintained by pluripotent stem cells. *Nature*, 439, 470.
- OHLSTEIN, B. & SPRADLING, A. 2007. Multipotent *Drosophila* intestinal stem cells specify daughter cell fates by differential notch signaling. *Science*, 315, 988-992.
- OHNEDA, K. & YAMAMOTO, M. 2002. Roles of hematopoietic transcription factors GATA-1 and GATA-2 in the development of red blood cell lineage. *Acta haematologica*, 108, 237-245.
- OKTAY, M. H., LEE, Y.-F., HARNEY, A., FARRELL, D., KUHN, N. Z., MORRIS, S. A., GREENSPAN, E., MOHLA, S., GRODZINSKI, P. & NORTON, L. 2015. Cell-to-cell communication in cancer: workshop report. *NPJ breast cancer*, 1, 15022.

- OKUMURA, T., MATSUMOTO, A., TANIMURA, T. & MURAKAMI, R. 2005. An endoderm-specific GATA factor gene, dGATAe, is required for the terminal differentiation of the *Drosophila* endoderm. *Dev Biol*, 278, 576-86.
- OKUMURA, T., TAJIRI, R., KOJIMA, T., SAIGO, K. & MURAKAMI, R. 2007. GATAe-dependent and -independent expressions of genes in the differentiated endodermal midgut of *Drosophila*. *Gene Expr Patterns*, 7, 178-86.
- OKUMURA, T., TAKEDA, K., KUCHIKI, M., AKAISHI, M., TANIGUCHI, K. & ADACHI-YAMADA, T. 2016. GATAe regulates intestinal stem cell maintenance and differentiation in *Drosophila* adult midgut. *Dev Biol*, 410, 24-35.
- PAGNI, M., IOANNIDIS, V., CERUTTI, L., ZAHN-ZABAL, M., JONGENEEL, C. V., HAU, J., MARTIN, O., KUZNETSOV, D. & FALQUET, L. 2007. MyHits: improvements to an interactive resource for analyzing protein sequences. *Nucleic acids research*, 35, W433-W437.
- PALANKER, L., TENNESSEN, J. M., LAM, G. & THUMMEL, C. S. 2009. *Drosophila* HNF4 regulates lipid mobilization and beta-oxidation. *Cell Metab*, 9, 228-39.
- PALU, R. A. & THUMMEL, C. S. 2016. Sir2 acts through Hepatocyte nuclear factor 4 to maintain insulin signaling and metabolic homeostasis in *Drosophila*. *PLoS genetics*, 12, e1005978.
- PARKER, L., PADILLA, M., DU, Y., DONG, K. & TANOUYE, M. A. 2011. *Drosophila* as a model for epilepsy: bss is a gain-of-function mutation in the para sodium channel gene that leads to seizures. *Genetics*, 187, 523-534.
- PARKHITKO, A. A., BINARI, R., ZHANG, N., ASARA, J. M., DEMONTIS, F. & PERRIMON, N. 2016. Tissue-specific down-regulation of S-adenosyl-homocysteine via suppression of dAhcyL1/dAhcyL2 extends health span and life span in *Drosophila*. *Genes Dev*, 30, 1409-22.
- PARVY, J.-P., HODGSON, J. A. & CORDERO, J. B. 2018. *Drosophila* as a Model System to Study Nonautonomous Mechanisms Affecting Tumour Growth and Cell Death. *BioMed research international*, 2018.
- PATIENT, R. K. & MCGHEE, J. D. 2002. The GATA family (vertebrates and invertebrates). *Curr Opin Genet Dev*, 12, 416-22.
- PÉREZ-GÓMEZ, R., SLOVÁKOVÁ, J., RIVES-QUINTO, N., KREJCI, A. & CARMENA, A. 2013. A Serrate-Notch-Canoe complex mediates essential interactions between glia and neuroepithelial cells during *Drosophila* optic lobe development. *J Cell Sci*, 126, 4873-4884.
- PEREZ-PINERA, P., KOCAK, D. D., VOCKLEY, C. M., ADLER, A. F., KABADI, A. M., POLSTEIN, L. R., THAKORE, P. I., GLASS, K. A., OUSTEROUT, D. G. & LEONG, K. W. 2013a. RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nature methods*, 10, 973.
- PEREZ-PINERA, P., OUSTEROUT, D. G., BRUNGER, J. M., FARIN, A. M., GLASS, K. A., GUILAK, F., CRAWFORD, G. E., HARTEMINK, A. J. & GERSBACH, C. A.

- 2013b. Synergistic and tunable human gene activation by combinations of synthetic transcription factors. *Nature methods*, 10, 239.
- PETERS, I., DUBROWINSKAJA, N., KOGOSOV, M., ABBAS, M., HENNENLOTTER, J., VON KLOT, C., MERSEBURGER, A. S., STENZL, A., SCHERER, R., KUCZYK, M. A. & SERTH, J. 2014a. Decreased GATA5 mRNA expression associates with CpG island methylation and shortened recurrence-free survival in clear cell renal cell carcinoma. *BMC Cancer*, 14, 101.
- PETERS, I., GEBAUER, K., DUBROWINSKAJA, N., ATSCHEKZEI, F., KRAMER, M. W., HENNENLOTTER, J., TEZVAL, H., ABBAS, M., SCHERER, R., MERSEBURGER, A. S., STENZL, A., KUCZYK, M. A. & SERTH, J. 2014b. GATA5 CpG island hypermethylation is an independent predictor for poor clinical outcome in renal cell carcinoma. *Oncol Rep*, 31, 1523-30.
- PFEIFFER, B. D., NGO, T.-T. B., HIBBARD, K. L., MURPHY, C., JENETT, A., TRUMAN, J. W. & RUBIN, G. M. 2010. Refinement of tools for targeted gene expression in *Drosophila*. *Genetics*.
- PIHLAJOKI, M., FÄRKILÄ, A., SOINI, T., HEIKINHEIMO, M. & WILSON, D. B. 2016. GATA factors in endocrine neoplasia. *Molecular and cellular endocrinology*, 421, 2-17.
- PISITKUN, T., HOFFERT, J. D., SAEED, F. & KNEPPER, M. A. 2011. NHLBI-AbDesigner: an online tool for design of peptide-directed antibodies. *American Journal of Physiology-Cell Physiology*, 302, C154-C164.
- PISITKUN, T., HOFFERT, J. D., SAEED, F. & KNEPPER, M. A. 2012. NHLBI-AbDesigner: an online tool for design of peptide-directed antibodies. *Am J Physiol Cell Physiol*, 302, C154-64.
- PISTON, D. W. 2005. When two is better than one: Elements of intravital microscopy. *PLoS biology*, 3, e207.
- RAFTERY, L. A. & SUTHERLAND, D. J. 2003. Gradients and thresholds: BMP response gradients unveiled in *Drosophila* embryos. *TRENDS in Genetics*, 19, 701-708.
- RAUSKOLB, C., PAN, G., REDDY, B., OH, H. & IRVINE, K. D. 2011. Zyxin links fat signaling to the hippo pathway. *PLoS biology*, 9, e1000624.
- REHORN, K. P., THELEN, H., MICHELSON, A. M. & REUTER, R. 1996. A molecular aspect of hematopoiesis and endoderm development common to vertebrates and *Drosophila*. *Development*, 122, 4023-4031.
- REN, C. Y., AKIYAMA, Y., MIYAKE, S. & YUASA, Y. 2004. Transcription factor GATA-5 selectively up-regulates mucin gene expression. *J Cancer Res Clin Oncol*, 130, 245-52.
- REN, F., WANG, B., YUE, T., YUN, E. Y., IP, Y. T. & JIANG, J. 2010. Hippo signaling regulates *Drosophila* intestine stem cell proliferation through multiple pathways. *Proc Natl Acad Sci U S A*, 107, 21064-9.

- REYA, T., MORRISON, S. J., CLARKE, M. F. & WEISSMAN, I. L. 2001. Stem cells, cancer, and cancer stem cells. *nature*, 414, 105-111.
- RICHARDSON, D. S. & LICHTMAN, J. W. 2015. Clarifying tissue clearing. *Cell*, 162, 246-257.
- RODRIGUES, N. P., BOYD, A. S., FUGAZZA, C., MAY, G. E., GUO, Y., TIPPING, A. J., SCADDEN, D. T., VYAS, P. & ENVER, T. 2008. GATA-2 regulates granulocyte-macrophage progenitor cell function. *Blood*, 112, 4862-4873.
- RODRIGUEZ-BRAVO, V., CARCELES-CORDON, M., HOSHIDA, Y., CORDON-CARDO, C., GALSKEY, M. D. & DOMINGO-DOMENECH, J. 2017. The role of GATA2 in lethal prostate cancer aggressiveness. *Nature Reviews Urology*, 14, 38.
- ROTHENPIELER, U. W. & DRESSLER, G. R. 1993. Pax-2 is required for mesenchyme-to-epithelium conversion during kidney development. *Development*, 119, 711-720.
- SALVAING, J., LOPEZ, A., BOIVIN, A., DEUTSCH, J. S. & PERONNET, F. 2003. The Drosophila Corto protein interacts with Polycomb-group proteins and the GAGA factor. *Nucleic Acids Res*, 31, 2873-82.
- SAM, S., LEISE, W. & HOSHIZAKI, D. K. 1996. The serpent gene is necessary for progression through the early stages of fat-body development. *Mechanisms of development*, 60, 197-205.
- SANTOS, M. G., JORGE, S. A., BRILLET, K. & PEREIRA, C. A. 2007. Improving heterologous protein expression in transfected Drosophila S2 cells as assessed by EGFP expression. *Cytotechnology*, 54, 15-24.
- SAXENA, A., DENHOLM, B., BUNT, S., BISCHOFF, M., VIJAYRAGHAVAN, K. & SKAER, H. 2014. Epidermal growth factor signalling controls myosin II planar polarity to orchestrate convergent extension movements during Drosophila tubulogenesis. *PLoS Biol*, 12, e1002013.
- SCHMITTGEN, T. D. & LIVAK, K. J. 2008. Analyzing real-time PCR data by the comparative C T method. *Nature protocols*, 3, 1101.
- SCHNEIDER, D. & SHAHABUDDIN, M. 2000. Malaria parasite development in a Drosophila model. *Science*, 288, 2376-2379.
- SENGER, K., HARRIS, K. & LEVINE, M. 2006. GATA factors participate in tissue-specific immune responses in Drosophila larvae. *Proc Natl Acad Sci U S A*, 103, 15957-62.
- SHARROCKS, A. D. 2001. The ETS-domain transcription factor family. *Nature reviews Molecular cell biology*, 2, 827.
- SHI, Q. & KING, R. W. 2005. Chromosome nondisjunction yields tetraploid rather than aneuploid cells in human cell lines. *Nature*, 437, 1038.
- SHIMIZU, R., TAKAHASHI, S., OHNEDA, K., ENGEL, J. D. & YAMAMOTO, M. 2001. In vivo requirements for GATA-1 functional domains during primitive and definitive erythropoiesis. *The EMBO journal*, 20, 5250-5260.

- SHOAIB, M. 2011. *Development of new approaches to study the role of chromatin in dna damage response*. Université Paris Sud-Paris XI.
- SHUREIQI, I., ZUO, X., BROADDUS, R., WU, Y., GUAN, B., MORRIS, J. S. & LIPPMAN, S. M. 2007. The transcription factor GATA-6 is overexpressed in vivo and contributes to silencing 15-LOX-1 in vitro in human colon cancer. *The FASEB Journal*, 21, 743-753.
- SINGH, S. R. & HOU, S. X. 2009. Multipotent stem cells in the Malpighian tubules of adult *Drosophila melanogaster*. *J Exp Biol*, 212, 413-23.
- SINGH, S. R., LIU, W. & HOU, S. X. 2007. The adult *Drosophila* malpighian tubules are maintained by multipotent stem cells. *Cell Stem Cell*, 1, 191-203.
- SINYUK, M., MULKEARNS-HUBERT, E. E., REIZES, O. & LATHIA, J. 2018. Cancer Connectors: Connexins, Gap Junctions, and Communication. *Frontiers in Oncology*, 8, 646.
- SKAER, H. 1989. Cell division in Malpighian tubule development in *D. melanogaster* is regulated by a single tip cell. *Nature*, 342, 566.
- SKAER, H. L. B., HARRISON, J. & MADDRELL, S. 1990. Physiological and structural maturation of a polarised epithelium: the Malpighian tubules of a blood-sucking insect, *Rhodnius prolixus*. *Journal of Cell Science*, 96, 537-547.
- SOMBKE, A., LIPKE, E., MICHALIK, P., UHL, G. & HARZSCH, S. 2015. Potential and limitations of X-Ray micro-computed tomography in arthropod neuroanatomy: A methodological and comparative survey. *Journal of Comparative Neurology*, 523, 1281-1295.
- SOUDAIS, C., BIELINSKA, M., HEIKINHEIMO, M., MACARTHUR, C. A., NARITA, N., SAFFITZ, J. E., SIMON, M. C., LEIDEN, J. M. & WILSON, D. B. 1995. Targeted mutagenesis of the transcription factor GATA-4 gene in mouse embryonic stem cells disrupts visceral endoderm differentiation in vitro. *Development*, 121, 3877-3888.
- SOUTHALL, T. D., GOLD, K. S., EGGER, B., DAVIDSON, C. M., CAYGILL, E. E., MARSHALL, O. J. & BRAND, A. H. 2013. Cell-type-specific profiling of gene expression and chromatin binding without cell isolation: assaying RNA Pol II occupancy in neural stem cells. *Developmental cell*, 26, 101-112.
- SOZEN, M. A., ARMSTRONG, J. D., YANG, M., KAISER, K. & DOW, J. A. 1997. Functional domains are specified to single-cell resolution in a *Drosophila* epithelium. *Proc Natl Acad Sci U S A*, 94, 5207-12.
- SÖZEN, M. A., ARMSTRONG, J. D., YANG, M., KAISER, K. & DOW, J. A. 1997. Functional domains are specified to single-cell resolution in a *Drosophila* epithelium. *Proceedings of the National Academy of Sciences*, 94, 5207-5212.
- SPAHN, P., HUELSMANN, S., REHORN, K.-P., MISCHKE, S., MAYER, M., CASALI, A. & REUTER, R. 2014. Multiple regulatory safeguards confine the expression of the GATA factor *Serpent* to the hemocyte primordium within the *Drosophila* mesoderm. *Developmental biology*, 386, 272-279.

- SPRADLING, A. C. & RUBIN, G. M. 1982. Transposition of cloned P elements into *Drosophila* germ line chromosomes. *Science*, 218, 341-347.
- STERGIOPOULOS, K., CABRERO, P., DAVIES, S.-A. & DOW, J. A. 2009. Salty dog, an SLC5 symporter, modulates *Drosophila* response to salt stress. *Physiological genomics*, 37, 1-11.
- STORCHOVA, Z. & PELLMAN, D. 2004. From polyploidy to aneuploidy, genome instability and cancer. *Nature reviews Molecular cell biology*, 5, 45.
- SUDARSAN, V., PASALODOS-SANCHEZ, S., WAN, S., GAMPEL, A. & SKAER, H. 2002. A genetic hierarchy establishes mitogenic signalling and mitotic competence in the renal tubules of *Drosophila*. *Development*, 129, 935-944.
- SUSAKI, E. A. & UEDA, H. R. 2016. Whole-body and whole-organ clearing and imaging techniques with single-cell resolution: toward organism-level systems biology in mammals. *Cell chemical biology*, 23, 137-157.
- SYED, Z. A., HÄRD, T., UV, A. & VAN DIJK-HÄRD, I. F. 2008. A potential role for *Drosophila* mucins in development and physiology. *PLoS One*, 3, e3041.
- TAINAKA, K., KUNO, A., KUBOTA, S. I., MURAKAMI, T. & UEDA, H. R. 2016. Chemical principles in tissue clearing and staining protocols for whole-body cell profiling. *Annual review of cell and developmental biology*, 32, 713-741.
- TAKASHIMA, S., PAUL, M., AGHAJANIAN, P., YOUNOSSI-HARTENSTEIN, A. & HARTENSTEIN, V. 2013. Migration of *Drosophila* intestinal stem cells across organ boundaries. *Development*, 140, 1903-11.
- TAKATSUJI, H. 1998. Zinc-finger transcription factors in plants. *Cellular and Molecular Life Sciences CMLS*, 54, 582-596.
- TARAVIRAS, S., MONAGHAN, A. P., SCHÜTZ, G. & KELSEY, G. 1994. Characterization of the mouse HNF-4 gene and its expression during mouse embryogenesis. *Mechanisms of development*, 48, 67-79.
- TERHZA, S., CABRERO, P., ROBBEN, J. H., RADFORD, J. C., HUDSON, B. D., MILLIGAN, G., DOW, J. A. & DAVIES, S.-A. 2012. Mechanism and function of *Drosophila* capa GPCR: a desiccation stress-responsive receptor with functional homology to human neuromedinU receptor. *PLoS One*, 7, e29897.
- TERHZA, S., FINLAYSON, A. J., STIRRAT, L., YANG, J., TRICOIRE, H., WOODS, D. J., DOW, J. A. & DAVIES, S. A. 2010. Cell-specific inositol 1,4,5 trisphosphate 3-kinase mediates epithelial cell apoptosis in response to oxidative stress in *Drosophila*. *Cell Signal*, 22, 737-48.
- TERHZA, S., O'CONNELL, F. C., POLLOCK, V. P., KEAN, L., DAVIES, S. A., VEENSTRA, J. A. & DOW, J. A. 1999. Isolation and characterization of a leucokinin-like peptide of *Drosophila melanogaster*. *Journal of Experimental Biology*, 202, 3667-3676.
- TERHZA, S., TEETS, N. M., CABRERO, P., HENDERSON, L., RITCHIE, M. G., NACHMAN, R. J., DOW, J. A., DENLINGER, D. L. & DAVIES, S.-A. 2015a.

Insect capa neuropeptides impact desiccation and cold tolerance. *Proceedings of the National Academy of Sciences*, 112, 2882-2887.

- TERHSAZ, S., TEETS, N. M., CABRERO, P., HENDERSON, L., RITCHIE, M. G., NACHMAN, R. J., DOW, J. A., DENLINGER, D. L. & DAVIES, S. A. 2015b. Insect capa neuropeptides impact desiccation and cold tolerance. *Proc Natl Acad Sci U S A*, 112, 2882-7.
- THOMAS, M. C. & CHIANG, C.-M. 2006. The general transcription machinery and general cofactors. *Critical reviews in biochemistry and molecular biology*, 41, 105-178.
- THOMAS, U., SPEICHER, S. A. & KNUST, E. 1991. The *Drosophila* gene *Serrate* encodes an EGF-like transmembrane protein with a complex expression pattern in embryos and wing discs. *Development*, 111, 749-761.
- TIAN, A., BENCHABANE, H., WANG, Z., ZIMMERMAN, C., XIN, N., PEROCHON, J., KALNA, G., SANSOM, O. J., CHENG, C., CORDERO, J. B. & AHMED, Y. 2017. Intestinal stem cell overproliferation resulting from inactivation of the APC tumor suppressor requires the transcription cofactors *Earthbound* and *Erect wing*. *PLoS Genet*, 13, e1006870.
- TIAN, Y., YUAN, L., GOSS, A. M., WANG, T., YANG, J., LEPORE, J. J., ZHOU, D., SCHWARTZ, R. J., PATEL, V. & COHEN, E. D. 2010. Characterization and in vivo pharmacological rescue of a Wnt2-Gata6 pathway required for cardiac inflow tract development. *Developmental cell*, 18, 275-287.
- TOKUSUMI, Y., TOKUSUMI, T. & SCHULZ, R. A. 2018. Mechanical stress to *Drosophila* larvae stimulates a cellular immune response through the JAK/STAT signaling pathway. *Biochemical and biophysical research communications*.
- TOLEDO-ORTIZ, G., HUQ, E. & QUAIL, P. H. 2003. The *Arabidopsis* basic/helix-loop-helix transcription factor family. *The Plant Cell*, 15, 1749-1770.
- TOOTLE, T. L. & REBAY, I. 2005. Post-translational modifications influence transcription factor activity: a view from the ETS superfamily. *Bioessays*, 27, 285-298.
- TRINH, I. & BOULIANNE, G. L. 2013. Modeling obesity and its associated disorders in *Drosophila*. *Physiology*, 28, 117-124.
- TSAI, F.-Y., KELLER, G., KUO, F. C., WEISS, M., CHEN, J., ROSENBLATT, M., ALT, F. W. & ORKIN, S. H. 1994. An early haematopoietic defect in mice lacking the transcription factor GATA-2. *Nature*, 371, 221.
- TULLET, J. M., HERTWECK, M., AN, J. H., BAKER, J., HWANG, J. Y., LIU, S., OLIVEIRA, R. P., BAUMEISTER, R. & BLACKWELL, T. K. 2008. Direct inhibition of the longevity-promoting factor SKN-1 by insulin-like signaling in *C. elegans*. *Cell*, 132, 1025-1038.
- UGUR, B., CHEN, K. & BELLEN, H. J. 2016. *Drosophila* tools and assays for the study of human diseases. *Disease models & mechanisms*, 9, 235-244.

- VAN HAMBURG, J. P., DE BRUIJN, M. J., DINGJAN, G. M., BEVERLOO, H. B., DIEPSTRATEN, H., LING, K.-W. & HENDRIKS, R. W. 2008. Cooperation of Gata3, c-Myc and Notch in malignant transformation of double positive thymocytes. *Molecular immunology*, 45, 3085-3095.
- VAN HIEL, M. B., VAN LOY, T., POELS, J., VANDERSMISSEN, H. P., VERLINDEN, H., BADISCO, L. & BROECK, J. V. 2010. Neuropeptide receptors as possible targets for development of insect pest control agents. *Neuropeptide systems as targets for parasite and pest control*. Springer.
- VIDAL, S. J., RODRIGUEZ-BRAVO, V., QUINN, S. A., RODRIGUEZ-BARRUECO, R., LUJAMBIO, A., WILLIAMS, E., SUN, X., DE LA IGLESIA-VICENTE, J., LEE, A. & READHEAD, B. 2015. A targetable GATA2-IGF2 axis confers aggressiveness in lethal prostate cancer. *Cancer cell*, 27, 223-239.
- VINSON, C. R., SIGLER, P. B. & MCKNIGHT, S. L. 1989. Scissors-grip model for DNA recognition by a family of leucine zipper proteins. *Science*, 246, 911-916.
- VONK, P. J. & OHM, R. A. 2018. The role of homeodomain transcription factors in fungal development. *Fungal Biology Reviews*.
- VORBRÜGGEN, G., CONSTIEN, R., ZILIAN, O., WIMMER, E. A., DOWE, G., TAUBERT, H., NOLL, M. & JÄCKLE, H. 1997. Embryonic expression and characterization of a Ptx1 homolog in *Drosophila*. *Mechanisms of development*, 68, 139-147.
- WALL, B. J., OSCHMAN, J. L. & SCHMIDT, B. A. 1975. Morphology and function of Malpighian tubules and associated structures in the cockroach, *Periplaneta americana*. *Journal of morphology*, 146, 265-306.
- WALLACE, B. & KING, J. C. 1951. Genetic changes in populations under irradiation. *The American Naturalist*, 85, 209-222.
- WALTZER, L., BATAILLÉ, L., PEYREFITTE, S. & HAENLIN, M. 2002. Two isoforms of Serpent containing either one or two GATA zinc fingers have different roles in *Drosophila* haematopoiesis. *The EMBO journal*, 21, 5477-5486.
- WANG, J., KEAN, L., YANG, J., ALLAN, A. K., DAVIES, S. A., HERZYK, P. & DOW, J. A. 2004. Function-informed transcriptome analysis of *Drosophila* renal tubule. *Genome Biol*, 5, R69.
- WANGLER, M. F., YAMAMOTO, S. & BELLEN, H. J. 2015. Fruit flies in biomedical research. *Genetics*, 199, 639-653.
- WATERHOUSE, A., BERTONI, M., BIENERT, S., STUDER, G., TAURIELLO, G., GUMIENNY, R., HEER, F. T., DE BEER, T. A. P., REMPFER, C. & BORDOLI, L. 2018. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic acids research*.
- WEAVERS, H., PRIETO-SÁNCHEZ, S., GRAWE, F., GARCIA-LÓPEZ, A., ARTERO, R., WILSCH-BRÄUNINGER, M., RUIZ-GÓMEZ, M., SKAER, H. & DENHOLM, B. 2009. The insect nephrocyte is a podocyte-like cell with a filtration slit diaphragm. *Nature*, 457, 322.

- WEAVERS, H. & SKAER, H. 2013. Tip cells act as dynamic cellular anchors in the morphogenesis of looped renal tubules in *Drosophila*. *Dev Cell*, 27, 331-44.
- WECHSLER, J., GREENE, M., MCDEVITT, M. A., ANASTASI, J., KARP, J. E., LE BEAU, M. M. & CRISPINO, J. D. 2002. Acquired mutations in GATA1 in the megakaryoblastic leukemia of Down syndrome. *Nature genetics*, 32, 148.
- WEIGELIN, B., BAKKER, G.-J. & FRIEDL, P. 2016. Third harmonic generation microscopy of cells and tissue organization. *J Cell Sci*, jcs. 152272.
- WEISS, M. J., KELLER, G. & ORKIN, S. H. 1994. Novel insights into erythroid development revealed through in vitro differentiation of GATA-1 embryonic stem cells. *Genes & development*, 8, 1184-1197.
- WESSING, A. & EICHELBERG, D. 1979. Malpighian tubules, rectal papillae and excretion. *Genetics and biology of Drosophila*.
- WHISSELL, G., MONTAGNI, E., MARTINELLI, P., HERNANDO-MOMBLONA, X., SEVILLANO, M., JUNG, P., CORTINA, C., CALON, A., ABULI, A. & CASTELLS, A. 2014. The transcription factor GATA6 enables self-renewal of colon adenoma stem cells by repressing BMP gene expression. *Nature Cell Biology*, 16, 695.
- WOODS, D. F. & BRYANT, P. J. 1991. The discs-large tumor suppressor gene of *Drosophila* encodes a guanylate kinase homolog localized at septate junctions. *Cell*, 66, 451-464.
- WU, D., SUNKEL, B., CHEN, Z., LIU, X., YE, Z., LI, Q., GRENADE, C., KE, J., ZHANG, C. & CHEN, H. 2014. Three-tiered role of the pioneer factor GATA2 in promoting androgen-dependent gene expression in prostate cancer. *Nucleic acids research*, 42, 3607-3622.
- WU, M., PASTOR-PAREJA, J. C. & XU, T. 2010. Interaction between Ras V12 and scribbled clones induces tumour growth and invasion. *Nature*, 463, 545.
- XIE, Y., LI, X., DENG, X., HOU, Y., O'HARA, K., URSO, A., PENG, Y., CHEN, L. & ZHU, S. 2016. The Ets protein Pointed prevents both premature differentiation and dedifferentiation of *Drosophila* intermediate neural progenitors. *Development*, dev. 137281.
- YANG, J., MCCART, C., WOODS, D. J., TERHSAZ, S., GREENWOOD, K. G., FFRENCH-CONSTANT, R. H. & DOW, J. A. 2007. A *Drosophila* systems approach to xenobiotic metabolism. *Physiological genomics*, 30, 223-231.
- YAPICI, N., KIM, Y.-J., RIBEIRO, C. & DICKSON, B. J. 2008. A receptor that mediates the post-mating switch in *Drosophila* reproductive behaviour. *Nature*, 451, 33.
- ZARIN, A. A., ASADZADEH, J., HOKAMP, K., MCCARTNEY, D., YANG, L., BASHAW, G. J. & LABRADOR, J.-P. 2014. A transcription factor network coordinates attraction, repulsion, and adhesion combinatorially to control motor axon pathway selection. *Neuron*, 81, 1297-1311.

- ZARIN, A. A., DALY, A. C., HÜLSMEIER, J., ASADZADEH, J. & LABRADOR, J.-P. 2012. A GATA/homeodomain transcriptional code regulates axon guidance through the Unc-5 receptor. *Development*, 139, 1798-1805.
- ZENG, X., SINGH, S. R., HOU, D. & HOU, S. X. 2010. Tumor suppressors Sav/Scrib and oncogene Ras regulate stem-cell transformation in adult *Drosophila* malpighian tubules. *J Cell Physiol*, 224, 766-74.
- ZHAI, Z., BOQUETE, J.-P. & LEMAITRE, B. 2018. Cell-Specific Imd-NF- κ B Responses Enable Simultaneous Antibacterial Immunity and Intestinal Epithelial Cell Shedding upon Bacterial Infection. *Immunity*, 48, 897-910. e7.
- ZHAI, Z., BOQUETE, J. P. & LEMAITRE, B. 2017. A genetic framework controlling the differentiation of intestinal stem cells during regeneration in *Drosophila*. *PLoS Genet*, 13, e1006854.
- ZHANG, P., JUDY, M., LEE, S.-J. & KENYON, C. 2013. Direct and indirect gene regulation by a life-extending FOXO protein in *C. elegans*: roles for GATA factors and lipid gene regulators. *Cell metabolism*, 17, 85-100.
- ZHENG, R. & BLOBEL, G. A. 2010. GATA transcription factors and cancer. *Genes & cancer*, 1, 1178-1188.
- ZOHAR-STOOPEL, A., GONEN, N., MAHROUM, M., BEN-ZVI, D. S., TOLEDANO, H. & SALZBERG, A. 2014. Homothorax plays autonomous and nonautonomous roles in proximodistal axis formation and migration of the *Drosophila* renal tubules. *Dev Dyn*, 243, 132-44.
- ZOHAR-STOOPEL, A., GONEN, N., MAHROUM, M., BEN-ZVI, D. S., TOLEDANO, H. & SALZBERG, A. 2014. Homothorax plays autonomous and nonautonomous roles in proximodistal axis formation and migration of the *Drosophila* renal tubules. *Developmental Dynamics*, 243, 132-144.